

## **MALE NORTHERN GRAY GIBBON SONGS HAVE A PHRASE-INSERTING STRUCTURE**

YOICHI INOUE<sup>\*1</sup>, WAIDI SINUN<sup>2</sup>, SHIGETO YOSIDA<sup>3</sup> and KAZUO OKANOYA<sup>4</sup>

<sup>\*</sup>Corresponding Author: gibbonyoichi@hotmail.co.jp

<sup>1</sup>Cognition and Behavior Joint Research Laboratory, Riken Center for Brain Science Institute, Wako, Japan

<sup>2</sup>Research and Development Division Yayasan Sabah Group, Kota Kinabalu, Malaysia

<sup>3</sup>IQVIA Solutions Japan K.K., Tokyo, Japan

<sup>4</sup>Advanced Comprehensive Research Organization, Teikyo University, Tokyo, Japan

### **1. Introduction**

To date, there have been no reports of a phrase-embedding structure in animal vocal communication, although there are several reports of combinational animal sounds (Hauser et al., 2014). Gibbons (Hylobatidae) living in South-East Asia are small apes, among the closest living relatives of humans. Gibbons are known for their remarkable vocal behavior. The songs of male Northern Gray gibbons (*Hylobates funereus*) in the Danum Valley Conservation Area (DVCA), Sabah, Malaysia consist of two notes, “wa” and “oo”, and combinations of these are flexible (Inoue et al., 2017). The “wa” notes were generally short, with a rapid rise in frequency, and the “oo” notes were relatively monotonous. The phrases of a song have various note orders, i.e., the notes could be combined in different phrase types. The Northern Gray gibbon is the species formerly named as Mueller’s gibbon (*Hylobates muelleri*).

When phrase “N” is inserted within another phrase “AB”, the generated phrase is shown as “ANB”. We named this structure a phrase-inserting structure. In male Northern Gray gibbon songs in our study area, a succession of “wa” notes (trill) are sometimes placed at both the start and end of phrases. 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 cooccurs 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.”

If the phrase “N” itself also has an inserting structure, then ANB is represented as AABB, which can be called an embedding structure, as defined by Abe and Watanabe (2011). When the animals sing songs by combining acoustic elements flexibly, a complicated syntax may emerge. We hypothesized that there are phrase-inserting structures in male Northern Gray gibbon songs, and recorded songs by a male in the DVCA for analysis. This report is based on a paper by Inoue et al. (2020).

## 2. Methods

All observations were conducted in the DVCA located in Sabah, Malaysia, in the northeast region of Borneo Island (Fig. 1). We studied a gibbon group named “SAPA”, whose territory was located around the Borneo Rainforest Lodge (BRL). With the aid of a field assistant, we conducted a survey over 4–7 successive days, biannually in August and December, from 2001 to 2009. During the study period from 2001 to 2009, we recorded 70 songs in 107 days of observations. 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 started recording gibbon songs, as soon as they started singing. We used a digital audio recorder (R-09; Roland, Hamamatsu, Japan) with a microphone (ATM57; Audio-Technica, Tokyo, Japan), and 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. We analyzed 8,046 phrases in 70 songs from the SAPA adult male. We converted the recorded sounds to sonograms using Avisoft-SAS Lab Pro software (Avisoft, Berlin, Germany). Focusing on the fundamental frequency, we performed the spectral and temporal measurements described below. To remove ambient noise, we processed the sound through a high-pass filter to cut off sound below 500 Hz. Finally, sonograms were created for on-screen measurements (settings: 256-point fast Fourier transformation and Hamming windows). 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). Most of the note intervals in our subject male’s songs were less than 2.0 seconds (90.7%). Therefore, we identified different phrases within a song as being separated by pauses of > 2 seconds.

In order to confirm two note types, we selected a 5 % random sample of all notes 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. Based on the measured data, we calculated the duration,  $\Delta$  frequency, and  $\Delta$  frequency/duration of each note. Two-tailed t tests were used to compare the acoustic characteristics of the two notes.

Furthermore, we analyzed whether or not the corresponding phrases were indistinguishable in terms of the acoustic similarity between those with an inserted phrase (A-N-B) and with no inserted phrase (A-B). We selected 20 patterns of phrases and calculated the acoustical similarity using Avisoft-CORRELATOR for every pair of same type phrases. We calculated the acoustical similarity between all pairs of “A” and “B” 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 to 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 both from the same class (within-class condition) or from different classes (between-class condition). We performed a linear mixed-effect model entering the cross-correlation coefficients as the response variable and the within/between-class condition as the explanatory variable (fixed-effect). We entered the phrase types (the above-mentioned 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 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.

### 3. Results

The mean values of the all examined acoustic parameters of notes were statistically different between the two notes (Table 1). In this study, three or more successive “wa” notes (trill) were placed at both the start and end of phrases in

2,726 (33.9%) of the 8,046 phrases. Among these, we identified 1,891 phrases (23.5%) as fixed phrases. Phrases consisting of three or more notes were sometimes included within fixed phrases (Fig. 2). We found 448 phrases (5.6%) that included phrases consisting of three 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 ( $X_1 = 33.15$ ,  $p < 0.01$ ).

#### 4. Discussion

We analyzed songs by a single male. There were no notes in the studied gibbon's songs with acoustic characteristics largely different from "wa" and "oo" notes. 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. 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.

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.

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.

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. This may be 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.

Table 1. Acoustic characteristics of “wa” and “oo” notes.

	Mean (SD)		<i>t</i>	<i>p</i>
	wa (n = 3518)	oo (n = 953)		
Start frequency (kHz)	0.797 (0.087)	0.712 (0.056)	45.49	<i>p</i> < 0.01
End frequency (kHz)	1.094 (0.167)	0.796 (0.081)	77.50	<i>p</i> < 0.01
Middle frequency (kHz)	0.932 (0.111)	0.751 (0.058)	68.17	<i>p</i> < 0.01
Δ frequency (kHz)	0.297 (0.146)	0.084 (0.074)	62.08	<i>p</i> < 0.01
Minimum frequency (kHz)	0.797 (0.087)	0.694 (0.035)	55.49	<i>p</i> < 0.01
Maximum frequency (kHz)	1.094 (0.167)	0.810 (0.079)	74.59	<i>p</i> < 0.01
Duration (second)	0.056 (0.027)	0.235 (0.087)	63.01	<i>p</i> < 0.01
Δ frequency / Duration (kHz / second)	6.211 (3.868)	0.414 (0.467)	86.58	<i>p</i> < 0.01

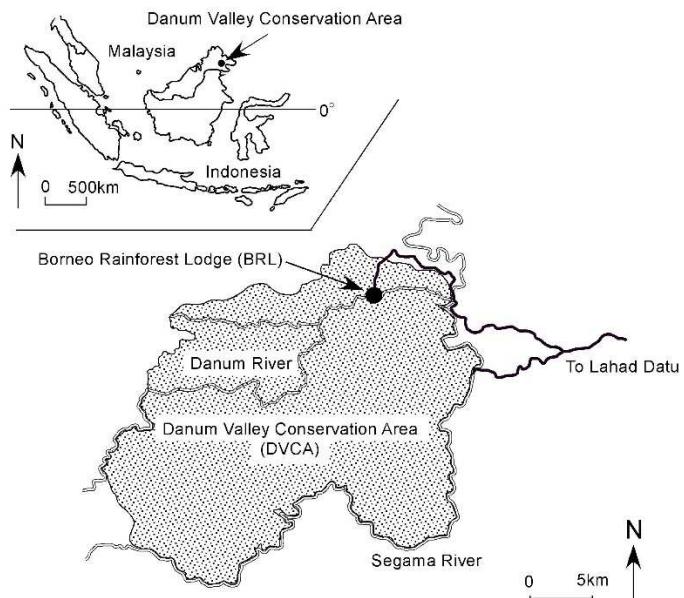


Figure 1. Location of the Borneo Rainforest Lodge (BRL) in the Danum Valley Conservation Area, Sabah, Malaysia.

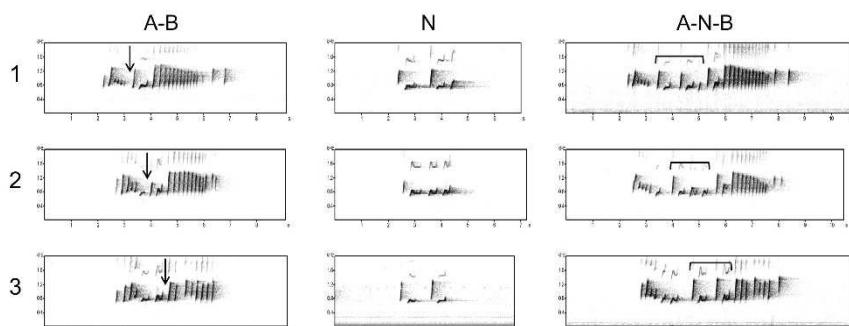


Figure 2. Examples of phrase A-N-B included within fixed phrases. Arrows indicate where phrase N was included within phrase A-B. Solid line brackets indicate phrase N included within phrase A-B.

## 5. References

- Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, 14, 1067–1074.

- Berwick, R. C., Beckers, G. J., Okanoya, K., & Bolhuis, J. J. (2012). A bird's eye view of human language evolution. *Frontiers in Evolutionary Neuroscience*, 4, 5.
- Chomsky, N. (2002). Syntactic structures, 9th edn. The Hague, The Netherlands: de Gruyter Mouton.
- Fitch, W. T. (2010). Three meanings of 'recursion': key distinctions for biolinguistics. In Larson, R. K., Deprez, V., & Yamakido, H. (eds): *The evolution of human language: Biolinguistic perspectives*, Cambridge: Cambridge University Press, pp. 73–90.
- Geissmann, T., Bohlen-Eyring, S., & Heuck, A. (2005). The male song of the Javan silvery gibbon (*Hylobates moloch*). *Contributions to Zoology*, 74, 1–25.
- Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., Chomsky, N., & Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in Psychology*, 5, 401.
- Inoue, Y., Sinun, W., Yosida, S., & Okanoya, K. (2017). Combinatory rules and chunk structure in male Mueller's gibbon songs. *Interaction Studies* 18, 1–25.
- Inoue, Y., Sinun, W., Yosida, S., & Okanoya, K. (2020). Note orders suggest phrase-inserting structure in male Mueller's gibbon songs: a case study. *acta ethologica*, 23(2), 89–102.
- Martins, M. D., & Fitch, W. T. (2014). Investigating recursion within a domain-specific framework. In *Language and recursion*. Springer, New York, NY. pp. 15–26.
- Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2019). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–143. URL: <https://CRAN.R-project.org/package=nlme>.
- Pinheiro, J. C., & Bates, D. M. (2000). "Mixed-Effects Models in S and S-PLUS", New York: Springer.
- Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, 95(2), 201–236.
- R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Thorpe, W. H. (1961). Bird-song. The biology of vocal communication and expression in birds. Cambridge monographs in experimental biology No. 12. Cambridge: University Press.