

Article

Temporal adjustment of short calls according to a partner during vocal turn-taking in Japanese macaques

Noriko KATSU^{a,b,c,*} Kazunori YAMADA^b, Kazuo OKANOYA^a, and Masayuki NAKAMICHI^b

^aGraduate School of Arts and Sciences, The University of Tokyo, Meguro, Tokyo, Japan, ^bGraduate School of Human Sciences, Osaka University, Suita, Osaka, Japan, and ^cJapan Society for the Promotion of Science, Tokyo, Japan

*Address correspondence to Noriko Katsu. E-mail: katsuno0414@gmail.com.

Handling editor: Andrea Ravignani

Received on 9 April 2018; accepted on 11 October 2018

Abstract

Turn-taking is a common feature in human speech, and is also seen in the communication of other primate species. However, evidence of turn-taking in vocal exchanges within a short time frame is still scarce in nonhuman primates. This study investigated whether dynamic adjustment during turn-taking in short calls exists in Japanese macaques *Macaca fuscata*. We observed exchanges of short calls such as grunts, girneys, and short, low coos during social interactions in a free-ranging group of Japanese macaques. We found that the median gap between the turns of two callers was 250 ms. Call intervals varied among individuals, suggesting that call intervals were not fixed among individuals. Solo call intervals were shorter than call intervals interrupted by responses from partners (i.e., exchanges) and longer than those between the partner's reply and the reply to that call, indicating that the monkeys did not just repeat calls at certain intervals irrespective of the social situation. The differences in call intervals during exchanged and solo call sequences were explained by the response interval of the partner, suggesting an adjustment of call timing according to the tempo of the partner's call utterance. These findings suggest that monkeys display dynamic temporal adjustment in a short time window, which is comparable with turn-taking in human speech.

Key words: turn-taking, rhythm, vocalization, primate.

Turn-taking, in which participants alternately reply to each other's utterances, occurs in human conversation, with a modal gap between turns of around 200 ms and the overlap between turns is typically minimized (Stivers et al. 2009; Levinson and Torreira 2015). To enable this rapid turn-taking, the speaker must anticipate the timing of the partner (Levinson and Torreira 2015). This system is universal across cultures, suggesting some biological basis, and possible phylogenetic continuity with other primate species (Levinson 2016).

Signal turn-taking is also found in many nonhuman animals in various taxa, such as insects, amphibians, birds, and mammals

(Zimmermann et al. 1989; Yoshida and Okanoya 2005; Takahashi et al. 2013; Henry et al. 2015; Ravignani 2015). The male advertisement signals (flash or sound) of many insects and frogs have species-specific patterns, and result in synchrony or alternation (i.e., turn-taking, Greenfield et al. 1997; Lewis and Cratsley 2008). As the term turn-taking refers to the exchange of communicative signals, such simultaneous signal production in nonhuman animals is also studied in the theoretical framework of "chorusing" (Kotz et al. 2018; Pika et al. 2018). The mechanism for this synchrony or alternation can be explained as a resettable oscillator. If females prefer leading signals, males may reset their signal rhythm upon hearing

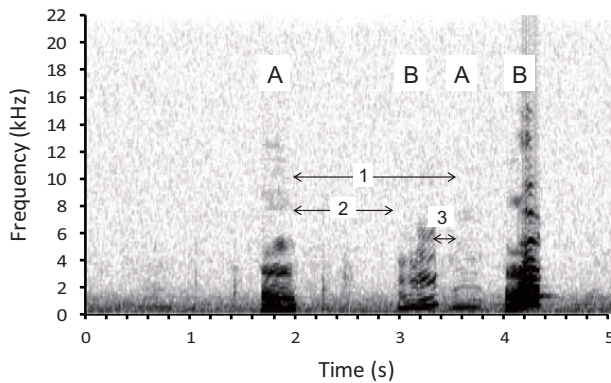


Figure 1. An example spectrogram of exchanged bouts of short calls (short low coo). The horizontal axis indicates time in seconds and the vertical axis indicates frequency in kHz. A focal individual (A) and a partner (B) alternately emitted short low coos. Arrow 1 indicates the phase response, Arrow 2 indicates the response interval, and Arrow 3 indicates the response-to-response interval.

their neighbors' signals and are known to restart their own periodic signal to avoid overlapping (Greenfield et al. 1997, 2016).

Vocal exchange in nonhuman primates is more likely to take the form of reciprocal call and response between two or more callers, rather than competitive interaction (Yoshida and Okanoya 2005). Pygmy marmosets *Cebuella pygmaea* tend to emit contact calls after hearing another's call, as opposed to the same individual calling consecutively (Snowdon and Cleveland 1984). In common marmosets *Callithrix jacchus*, the interval to reply to another's call was shorter than the interval to repeat consecutive calls when there was no reply (Miller and Wang 2006). This suggests that marmosets repeat their calls when there is no reply from others for a certain period. This difference between reply intervals and intervals between consecutive calls from the same caller without a reply has also been observed in other primate species (Masataka and Biben 1987; Sugiura 1993; Lemasson et al. 2011). A recent study revealed that common marmosets not only wait for a reply, but also dynamically adjust their call interval according to the reply latency of other callers (Takahashi et al. 2013). This finding suggests that turn-taking in marmosets displays the same characteristics as human speech, which can be explained by the model of coupled oscillators (Wilson and Wilson 2005).

Turn-taking in marmosets occurs in a relatively slow time window (gap within 5 s, Miller and Wang 2006; Takahashi et al. 2013) compared with that in human conversation (gap around 200 ms, Stivers et al. 2009). There is little information on turn-taking in nonhuman primates on a short time scale (but see Richman 1976), although short vocalizations are used in many species (Cheney and Seyfarth 1982; de Waal 1988; Bauers 1993; Whitham et al. 2007). Short, soft vocalizations are used by Japanese macaques *Macaca fuscata* in non-agonistic, face-to-face situations, and exchanged between interacting partners (Katsu et al. 2016). These short calls consist of three acoustically distinct call types, "grunts," "girneys," and "short low coos" (Figure 1). Girneys are tonal vocalizations, whereas grunts are atonal ones (Katsu et al. 2016). Short low coos are a type of coo call (Green 1975) that are shorter and lower in amplitude than other coo calls. These three vocalizations have the function of communicating non-agonistic intent, and affiliative interactions follow the calls with high probability irrespective of the call type when vocal exchanges occur (Katsu et al. 2016). Japanese macaques display both timing- and frequency-matching during vocal exchanges of contact calls (coo calls) and adjust call intervals

according to the distance of other group members (Sugiura 1998, 2007; Koda 2004). In experimental studies, rhesus macaques *Macaca mulatta*, which are closely related to the Japanese macaques, demonstrated an ability to detect rhythmic groups and produce regular rhythms in sub-second time windows (Zarco et al. 2009; Honing et al. 2012). This ability is also expected to be displayed during natural vocal interactions.

We aimed to clarify whether dynamic adjustment of call timing exists during vocal turn-taking in Japanese macaques. We first investigated whether call intervals were fixed among individuals during spontaneous call sequences, similar to the species-specific flashing patterns found in fireflies (Lampyridae, Lewis and Cratsley 2008). We thus predicted that if the call timing of an individual monkey is affected by the response from a partner during vocal exchange, rather than being independent of each other, the call intervals during vocal exchanges would be longer than those in the solo call sequences. We then examined the possibility that monkeys simply repeat calls at a certain fixed interval during vocal exchanges, or reset a call sequence upon hearing another's call replacing the first call of the call sequence of the initial caller with the recipient's call (resetting model). This model predicts that the interval between another individual's reply to the initial caller and the reply to this call from the initial caller fall in the range of the spontaneous calling tempo of the caller. Thus, we predicted that if a monkey did not just reset the phase of a call sequence, the intervals between the reply of a partner and the next subject's reply call would be shorter than the call intervals during solo call sequences. The reply latency to the other participant's response was expected to be shorter than the spontaneous calling tempo because when the monkeys actually interact with vocalizations, they reply before the initial caller repeats the call (Masataka and Biben 1987; Sugiura 1993; Miller and Wang 2006; Lemasson et al. 2011). We then investigated whether monkeys adjust their call timing according to the reply latency of their partners. We examined whether the differences in the duration of call intervals during call exchanges and solo calls were explained by the response latency of the partners. We predicted that, if they adjusted their call timing, the call intervals of an individual during exchanges would be extended according to the length of their partner's reply latency, compared with their call intervals during solo call sequences.

Materials and Methods

Study site and subjects

This study was conducted on a free-ranging group of Japanese macaques (Arashiyama group) at the Iwatayama Monkey Park (35°00'N and 135°67'E), Kyoto, Japan. Maternal kinships within this group have been recorded since the 1950s (Fedigan and Asquith 1991). The park staff fed the group daily with wheat grain and soybeans, 4 times at the provisioning ground and once near the sleeping site. We collected data between April and October 2012. The group included 126 individuals (88 adult [>4 years] females, 10 adult males, 10 immature females, and 18 immature males) at the beginning of the study. The average age of these adult females was 18.1 ± 8.4 years old (range: 4–33). The focal subjects were 15 adult females, who were on average 15.1 years old (range: 8–26). We chose full adult females because the usage of short calls differs between subadult females (4–5 years) and full adult females (Katsu et al. 2017), and males use short calls less frequently (unpublished data). We recorded the interactions between these focal females and any other adult females. The subjects were not maternally related to each other.

Data collection

Behavioral observations were conducted between 09:00 h and 17:30 h, when the monkeys remained near the provisioning area. The subjects were observed in a predetermined, randomized order. Dyadic social interactions of focal subjects were video-recorded (HDR-CX560V, SONY, Tokyo, Japan) by the observer using all occurrence sampling method during 20-min focal observations. During this 20-min observation, we recorded short calls, grunts, girneys, and coos using a directional microphone (NTG2, RODE, Sydney, Australia) and a digital audio recorder (V803, OLYMPUS, Tokyo, Japan), with a sampling rate of 44.1 kHz and 16-bit resolution. These vocalizations are distinct and were distinguishable by ear (Whitham et al. 2007; Katsu et al. 2016). We analyzed all short calls that a monkey emitted with its face oriented towards another monkey within 5 m (a recipient), as short calls were low in amplitude and could not be heard clearly from farther distances. We defined a call bout as a series of more than one call being emitted in <5 s (Whitham et al. 2007; Katsu et al. 2014). A vocal exchange was defined as a subject and a recipient both emitted calls within 5 s. The caller of each vocalization was determined by mouth movement and annotated orally in video recording. A total of 4,650 min of focal observation and 148 bouts of short calls were recorded. Out of these bouts, 84 consisted of the subjects' calls only (solo bouts), and 64 included at least one vocal exchange (exchanged bouts). We recorded on average 5.6 (SD: 3.5) solo bouts and 4.3 (SD: 3.1) exchanged bouts per subject. Moreover, 93 out of the 148 bouts consisted of different types of calls (e.g., a bout including both girneys and grunts) and the initial and response call type often differed.

Data analyses

Definition of call interval

Spectrograms were generated by fast Fourier transformation (FFT length = 512, overlap = 50%, sampling frequency = 22.05 kHz, with a Hamming window) using SASLab Pro software (Avisoft Bioacoustics, Glienicke, Germany). Call duration and call interval were calculated by 'automatic parameter measurement' within the software and inspected visually to an accuracy of 10 ms. The interval between calls was defined as the duration from the offset of the previous call to the onset of the subsequent call. Solo interval was defined as the duration from a call offset to the onset of the subsequent call produced by the same focal subject during solo bouts. Spontaneous calling tempo (T0) was defined as the median duration of the solo interval of each focal subject (Takahashi et al. 2013). We categorized call intervals during exchanged bouts following the methods of Greenfield et al. (1997) and Takahashi et al. (2013). "Phase response" (PR) was defined as a call interval of the same focal subject interrupted by the reply of a partner during an exchanged bout (Figure 1, arrow 1). Response interval (RI) was defined as the interval from the offset of a subject's call to the onset of the reply from a recipient (Figure 1, arrow 2). The response-to-response interval was defined as the interval between the offset of the partner's reply to the onset of the subject's next call and therefore the subjects reply to that reply (Figure 1, arrow 3). Most of the exchanged bouts included more than two-phase responses, that is, the first call of a certain phase response was sometimes the response to the preceding partner's call.

Statistics. We conducted permutation tests and linear mixed model (LMM) analyses in R version 3.2.0 (R Core Team 2013). LMMs were performed using the lme4 package (Bates et al. 2013) and then

a likelihood ratio test was performed with null models containing only random effects to examine the significance of the fixed, explanatory variable. We included random slopes for all fixed effects where applicable. All tests were two-tailed with significance levels set at $P < 0.05$.

Call duration. We compared the call durations of the subjects between solo and exchanged bouts to confirm that call duration was not affected by the occurrence of vocal exchange. To test this, we conducted LMM. The response variable was the duration of each call, and the explanatory variable was the type of the call, that is, whether it was emitted in a solo or exchanged bout. The random effects were subject ID and call bout ID. To compare the call duration between the subject and the partner, we also conducted LMM on the call duration during exchanged bouts, with the explanatory variable being whether the caller was a subject or a partner, and the random effects being subject ID and call bout ID.

Differences in call intervals between solo and exchanged bouts. We investigated whether the call interval is fixed among monkeys, or whether it is fixed between exchanged and solo bouts. We compared the call interval during solo bouts and the phase responses using LMM. The response variable was call interval, and the fixed effect was whether an interval was a solo interval or a phase response. Subject ID was included as a random effect. In this model, we also investigated whether there are individual differences between call intervals by comparing the full model and a model without the random effects of subject ID. This comparison was conducted to examine whether mean intervals varied between subjects (Kliegl et al. 2011). Solo and exchanged bouts that contained less than two calls from the subjects (simple call-and-response sequence or turn-taking initiated by a partner) were excluded from this analysis and those described below.

We then investigated the possibility that individuals repeated the fixed interval at the end of the response (resetting model). If these hypotheses were true, the response-to-response intervals would not differ from the solo intervals. We compared these two intervals using a permutation test. We excluded response-to-response intervals overlapping with the previous calls (interval < 0 ms). This is because solo intervals, by definition, do not take values < 0, and this might cause mean response-to-response intervals to be shorter than solo intervals.

Adjustment of call timing. We then examined whether call timing was adjusted according to the reply timing of the partner in exchanged bouts. The difference between phase response and T0 (PR - T0) indicates the amount of adjustment according to the reply. We predicted that if a subject did not lengthen or shorten the interval after their partner replied to their previous call, the phase response would be equal to T0 (PR - T0 = 0). If an individual adjusted the interval by a fixed amount or an amount unrelated to the reply latency of the partner (response interval: RI), there would be no correlation between PR - T0 and RI. To test this, we conducted an LMM and likelihood ratio test. The response variable was PR - T0, and the explanatory variable was RI corresponding to the PR. The random effects were subject ID and partner ID. We conducted bias correction following the method in Phoka et al. (2010), due to the non-uniformity of RI, and its potential to create a false positive when analyzing the relationship between phase response and response interval. The analysis was thus conducted with data including bias-correction data points.

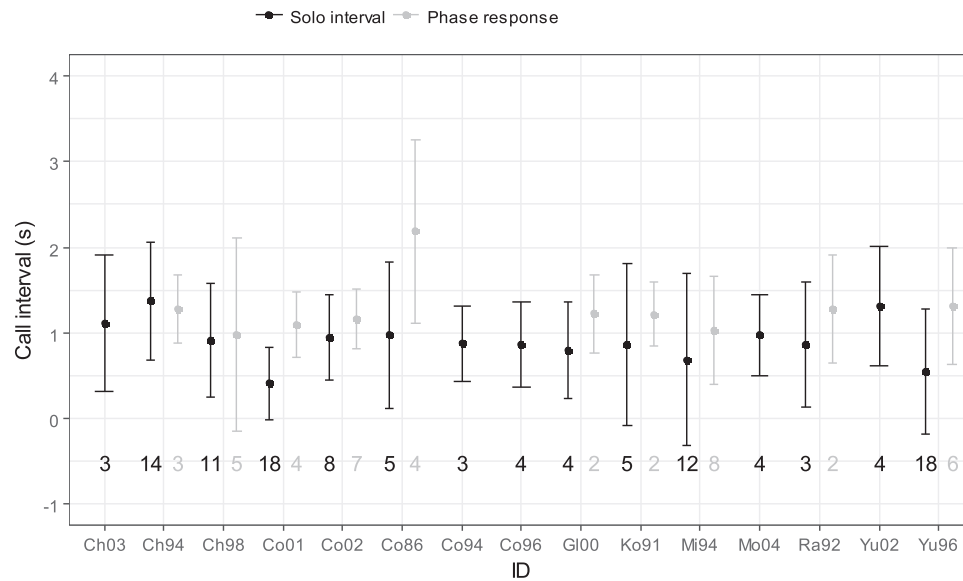


Figure 2. The mean and SD of solo intervals (black) and phase responses (gray) for each focal subject. Phase responses were not recorded in five of the 15 subjects. The horizontal axis indicates the ID of the subjects and the vertical axis indicates the duration in seconds (s). The numbers below the data point indicates the number of call intervals recorded for each subject.

Results

Number of calls and their duration

The median number of calls during one bout was 5.0 (range: 2–16) for exchanged bouts and 2.5 (2–23) for solo bouts. The median duration of one bout (from the onset of the first call to the offset of the last call of one bout) was 4,470 ms (range: 530–15,080 ms) for exchanged bouts and 2,800 ms (500–12,050 ms) for solo bouts. The median call duration of solo calls was 180 ms ($n=181$, range: 30–980 ms), and for exchanged calls the median duration for focal subjects was 200 ms ($n=109$, range: 60–530 ms). There were no significant differences in call duration between solo and exchanged bouts ($N=290$, LMM and likelihood ratio test, $\chi^2=1.968$, $df=1$, $P=0.161$, Supplementary Table S1), as well as between the call duration of subjects and partners ($N=157$, $\chi^2=1.410$, $df=1$, $P=0.235$, Supplementary Table S2).

Individual differences in call interval

Figure 2 illustrates the average and SD of solo intervals and phase responses by each subject. The median solo interval duration was 760 ms ($N=116$ from 41 bouts, range: 10–4,900 ms), and the median phase response duration was 1,220 ms ($N=43$ from 24 bouts, range: 140–2,330 ms). The LMM revealed that the type of interval had a significant effect on the duration of call intervals ($N=159$, Table 1), that is, phase responses were significantly longer than solo intervals. This finding supported predictions that monkeys do not simply emit calls at the same intervals during solo and exchanged bouts. The comparison between the full model and the model without the random effects of subject ID revealed that subject ID had a significant effect ($\chi^2=11.133$, $df=3$, $P=0.011$). This finding indicates that call intervals varied between individuals across the whole dataset.

Adjustment of call timing

The median call interval between two individuals in exchanged bouts was 250 ms ($N=106$, range: –120–4,790 ms). Call

overlapping between two individuals was observed in 6 out of 106 intervals (6%) in exchanged bouts. The median overlap duration of calls was 20 ms (10–120 ms).

The median response-to-response interval duration was 260 ms ($N=61$, range: –40–4,680 ms, including overlapping; Figure 3). Interval duration was significantly different between solo intervals and response-to-response intervals ($N=175$, excluding 2 response-to-response intervals <0 ms, permutation test: $P<0.002$, estimated mean difference: 520 ms). Response-to-response intervals were significantly shorter than solo intervals, supporting the predictions that monkeys do not repeat the same interval after the response of the partner.

The median duration of the response interval was 270 ms ($N=36$, range: 0–2,210 ms). We conducted analyses on whether intervals interrupted by responses were adjusted according to the response latency at the partner's end. An LMM and likelihood ratio test revealed that the difference between the call intervals of solo (TO) and exchanged (PR) bouts were significantly explained by the response interval (RI) of the partners ($N=68$, 7 individuals, LMM and likelihood ratio test: $\chi^2=5.175$, $df=1$, $P=0.023$, adjusted $R^2=0.389$; Figure 4, Table 2). This result suggests that the monkeys lengthen their reply intervals as their partner's reply latency increases.

Discussion

Turn-taking is important in terms of effective signal transmission (Yoshida and Okanoya 2005). We investigated whether dynamic adjustment was involved in turn-taking in the short calls of Japanese macaques, and our findings supported the idea that these monkeys coordinate their call timing during vocal exchanges.

Call intervals varied between individuals, suggesting that monkeys do not emit calls in a species-specific predetermined interval in spontaneous calls, as reported in fireflies (Lewis and Cratsley 2008). The differences in the median solo interval of different individuals could be as long as 1,500 ms. This could cause call overlapping between individuals if they emit calls at similar intervals to solo bouts during exchanged bouts. Therefore, some adjustment is needed to avoid

Table 1. Results of the linear mixed model showing the effect of type of interval on the call interval duration

Variables				
Random effects				
Groups		Variance	SD	Correlation
Subject ID	(Intercept)	0.077	0.278	
	Type of interval: phase response	0.057	0.239	−0.75
Residual		0.477	0.691	
Fixed effect		β (SE)	<i>t</i>	<i>P</i>
	(Intercept)	0.888 (0.109)	8.116	<0.0001
	Type of interval: phase response	0.389 (0.154)	2.532	0.043

The full versus null model comparison: $N = 159$, $\chi^2_3 = 11.728$, $P = 0.008$

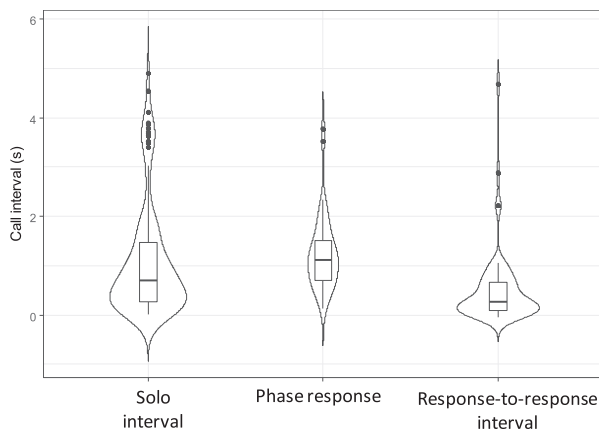


Figure 3. Violin plots and boxplots of call intervals for solo intervals, phase response, and response-to-response intervals. The horizontal axis indicates the type of call interval, and vertical axis indicates duration in seconds (s). Solo interval ($N = 166$): call intervals between calls emitted by a single individual with no partner; phase response ($N = 43$): call intervals between calls of the same subject interrupted by a reply in exchanged bouts; response-to-response interval ($N = 61$): intervals from the offset of a reply to the onset of the next call of the subject in exchanged bouts. Solo intervals were presented for comparison. Data from all subjects were pooled in this figure.

overlapping during vocal exchange. Call duration itself did not differ between solo and exchanged bouts, or between subjects and partners. This negates the possibility that calls would not overlap merely because call durations during exchanged bouts were too short.

The duration of phase responses, the call interval between two calls of the subject when interrupted by a reply, was significantly longer than the call intervals during solo bouts. Moreover, the resetting model, which involves resetting the calling tempo after hearing another's call (Greenfield et al. 1997) did not explain these data. These findings indicate that monkeys do not repeat calls at predetermined intervals regardless of the social situation, but reacted on their partner's reply. A previous study on Japanese macaques reported that the intervals of coo calls (contact calls) differed depending on the presence or absence of a reply (Sugiura 1993). This study also revealed the existence of adjustment of call timing in a time window (less than 1 s) shorter than that of coo calls; that is, subjects replied to a partner's reply in shorter latency than an interval that the same individual repeats the calls.

The difference, in the same individuals, between their phase responses during exchanged bouts and their intervals during solo bouts was explained by reply latency; that is, the call interval was extended according to the reply latency during vocal exchanges.

Some nonhuman primates have been known to wait for replies during vocal exchanges (Snowdon and Cleveland 1984; Masataka and Biben 1987; Sugiura 2007; Inoue et al. 2013). However, this dynamic temporal coordination between two callers has not been reported except in a study on marmosets (Takahashi et al. 2013). High flexibility in vocal production and usage, and developmental changes are exhibited by Japanese macaques (Koda 2004; Tanaka et al. 2006; Sugiura 2007; Katsu et al. 2017) as well as marmosets (Choi et al. 2015; Chow et al. 2015; Takahashi et al. 2016, 2017). Such flexibility in vocal production and usage may be the basis of temporal adjustment during turn-taking.

The modal gap during turn-taking is around 200 ms in human speech (Levinson and Torreira 2015), although physiological preparation before output takes 600–1,500 ms (Bates et al. 2003; Indefrey and Levelt 2004; Roberts et al. 2015). These studies indicate that timing prediction is involved in human speech. Turn-taking in Japanese macaques also occurred in a short time period (250 ms); however, this vocal production can be explained by reactive response. A periodic coupling in the vocal turn-taking was shown in humans (Wilson and Wilson 2005) and marmosets (Takahashi et al. 2013). Call sequences of Japanese macaques in our study included 5 calls on average, thus such periodic coupling might be less likely to exist. Nonetheless, it is worth further investigating the temporal signatures of call preparation and production to determine whether such timing prediction is involved in vocal turn-taking in Japanese macaques.

Affiliative interactions were more likely to occur when vocal exchange occurred (Katsu et al. 2016), however, whether the timing of a reply affects subsequent social interactions remains unclear. A previous study found that juveniles' reply to contact calls of others was less frequent than adults in Japanese macaques (Lemasson et al. 2013). In infant marmosets, turn-taking with parents gradually develops in the first year of life (Takahashi et al. 2016). The examination of the following social interactions and developmental changes in such species may reveal how timing adjustment during vocal turn-taking functions in their social life.

Acknowledgments

We are grateful for the kind support provided by Mr S. Asaba and the staff of the Iwatayama Monkey Park.

Funding

This study was supported by JSPS KAKENHI Grant Number 13J03612, 16H06928, and 17J10994, and MEXT/JSPS KAKENHI Grant Number #4903, JP17H06380.

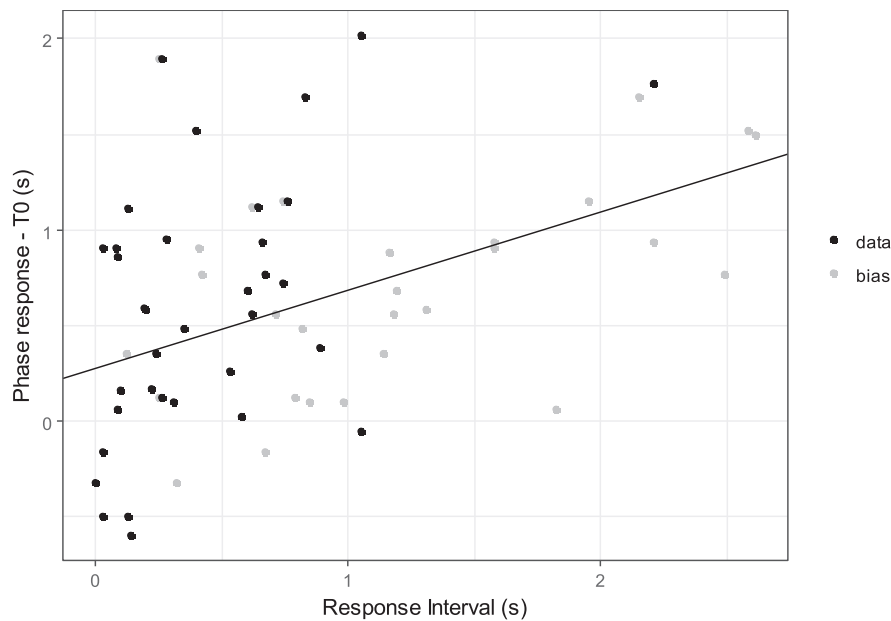


Figure 4. Relationship between response interval and the amount call intervals was adjusted by each subject. Horizontal axis indicates duration of response interval (RI) in seconds (s), and vertical axis indicates differences between phase response and call interval during solo bouts (PR – T0) in seconds (s). Bias correction points are represented by gray dots. The positive relationship indicates that two callers lengthened (or shortened) their response latency according to the length of their partner's response latency.

Table 2. Results of the linear mixed model showing the effect of response interval on the value of corresponding phase response - T0

Variables	β (SE)	t	P
Fixed effect			
(Intercept)	0.275 (0.235)	1.173	0.293
Response interval	0.408 (0.148)	2.758	0.038

The full versus null model comparison: $N=68$, $\chi^2_1=5.175$ $P=0.0229$, $R^2=0.389$.

Ethics statement

We conducted this study in accordance with the Regulations on Animal Experimentation at Osaka University. The study was approved by the Animal Research Committee of the Graduate School of Human Sciences at Osaka University in Japan (No. 21–14–10).

References

- Bates D, Maechler M, Bolker B, Walker S, 2013. lme4: linear mixed-effects models using Eigen and S4. *R Package Version 1.0–5*. Retrieved from <https://CRAN.R-project.org/package=lme4> (accessed 30 October 2018).
- Bates E, D'Amico S, Jacobsen T, Székely A, Andonova E, Devescovi A, Herron D, Lu CC, Pechmann T, Pléh C, Wicha N, Federmeier K, Gerdjikova I, Gutierrez G, Hung D, Hsu J, Iyer G, Kohnert K, Mehotcheva T, Orozco-Figueroa A, Tzeng A, Tzeng O, 2003. Timed picture naming in seven languages. *Psychon B Rev* 10:344–380.
- Bauers KA, 1993. A functional-analysis of staccato grunt vocalizations in the stump-tailed macaque *Macaca arctoides*. *Ethology* 94:147–161.
- Cheney DL, Seyfarth RM, 1982. How vervet monkeys perceive their grunts: field playback experiments. *Anim Behav* 30:739–751.
- Choi JY, Takahashi DY, Ghazanfar AA, 2015. Cooperative vocal control in marmoset monkeys via vocal feedback. *J Neurophysiol* 114:274–283.
- Chow CP, Mitchell JF, Miller CT, 2015. Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc R Soc B* 282:20150069.

- de Waal FBM, 1988. The communicative repertoire of captive bonobos *Pan paniscus*, compared to that of chimpanzees. *Behaviour* 108: 183–251.
- Fedigan LM, Asquith PJ, 1991. *The Monkeys of Arashiyama: Thirty-Five Years of Research in Japan and the West*. Albany (NY): State University of New York Press.
- Green SM, 1975. Variation of vocal pattern with social situation in the Japanese monkey *Macaca fuscata*: a field study. In: Rosenblum L, editor. *Primate Behavior*. New York: Academic Press. 1–102.
- Greenfield MD, Esquer-Garrigos Y, Streiff R, Party V, 2016. Animal choruses emerge from receiver psychology. *Sci Rep* 6:34369.
- Greenfield MD, Tourtellot MK, Snedden WA, 1997. Precedence effects and the evolution of chorusing. *Proc R Soc B* 264:1355–1361.
- Henry L, Craig AJ, Lemasson A, Hausberger M, 2015. Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model. *Front Psychol* 6:1416.
- Honing H, Merchant H, Håden GP, Prado L, Bartolo R, 2012. Rhesus monkeys *Macaca mulatta* detect rhythmic groups in music, but not the beat. *PLoS ONE* 7:e51369.
- Indefrey P, Levelt WJ, 2004. The spatial and temporal signatures of word production components. *Cognition* 92:101–144.
- 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.
- Katsu N, Yamada K, Nakamichi M, 2014. Development in the usage and comprehension of greeting calls in a free-ranging group of Japanese macaques *Macaca fuscata*. *Ethology* 120:1024–1034.
- Katsu N, Yamada K, Nakamichi M, 2016. Function of grunts, girneys and coo calls of Japanese macaques *Macaca fuscata* in relation to call usage, age and dominance relationships. *Behaviour* 153:125–142.
- Katsu N, Yamada K, Nakamichi M, 2017. Influence of social interactions with nonmother females on the development of call usage in Japanese macaques. *Anim Behav* 123:267–276.
- Kliegl R, Wei P, Dambacher M, Yan M, Zhou X, 2011. Experimental effects and individual differences in linear mixed models: estimating the relationship between spatial, object, and attraction effects in visual attention. *Front Psychol* 1:238.

- Koda H, 2004. Flexibility and context-sensitivity during the vocal exchange of coo calls in wild Japanese macaques *Macaca fuscata yakui*. *Behaviour* 141: 1279–1296.
- Kotz SA, Ravignani A, Fitch WT, 2018. The evolution of rhythm processing. *Trends Cogn Sci* 22:896–910.
- Lemasson A, Glas L, Barbu S, Lacroix A, Guilloux M et al., 2011. Youngsters do not pay attention to conversational rules: is this so for nonhuman primates? *Sci Rep* 1:22.
- Lemasson A, Guilloux M, Rizaldi Barbu S, Lacroix A et al., 2013. Age- and sex-dependent contact call usage in Japanese macaques. *Primates* 54:283–291.
- Levinson SC, 2016. Turn-taking in human communication—origins and implications for language processing. *Trends Cogn Sci* 20:6–14.
- Levinson SC, Torreira F, 2015. Timing in turn-taking and its implications for processing models of language. *Front Psychol* 6:731.
- Lewis SM, Cratsley CK, 2008. Flash signal evolution, mate choice, and predation in fireflies. *Annu Rev Entomol* 53:293–321.
- Masataka N, Biben M, 1987. Temporal rules regulating affiliative vocal exchanges of squirrel monkeys. *Behaviour* 101:311–319.
- Miller CT, Wang XQ, 2006. Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 192:27–38.
- Phoka E, Cuntz H, Roth A, Häusser M, 2010. A new approach for determining phase response curves reveals that Purkinje cells can act as perfect integrators. *PLoS Comput Biol* 6:e1000768.
- Pika S, Wilkinson R, Kendrick KH, Vernes SC, 2018. Taking turns: bridging the gap between human and animal communication. *Proc R Soc B* 285:20180598.
- Ravignani A, 2015. Evolving perceptual biases for antisynchrony: a form of temporal coordination beyond synchrony. *Front Neurol* 9:339.
- R Core Team, 2013. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. Retrieved from <http://www.R-project.org/> (accessed 30 October 2018).
- Richman B, 1976. Some vocal distinctive features used by gelada monkeys. *J Acoust Soc Am* 60:718–724.
- Roberts SG, Torreira F, Levinson SC, 2015. The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Front Psychol* 6:509.
- Snowdon CT, Cleveland J, 1984. Conversations among pygmy marmosets. *Am J Primatol* 7:15–20.
- Stivers T, Enfield N, Brown P, Englert C, Hayashi M et al., 2009. Universals and cultural variation in turn-taking in conversation. *Proc Natl Acad Sci U S A* 106:10587–10592.
- Sugiura H, 1993. Temporal and acoustic correlates in vocal exchange of coo calls in Japanese macaques. *Behaviour* 124:207–225.
- Sugiura H, 1998. Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Anim Behav* 55:673–687.
- Sugiura H, 2007. Adjustment of temporal call usage during vocal exchange of coo calls in Japanese macaques. *Ethology* 113:528–533.
- Takahashi DY, Narayanan DZ, Ghazanfar AA, 2013. Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr Biol* 23:2162–2168.
- Takahashi DY, Fenley AR, Ghazanfar AA, 2016. Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philos Trans R Soc Lond B Biol Sci* 371:20150370.
- Takahashi DY, Liao DA, Ghazanfar AA, 2017. Vocal learning via social reinforcement by infant marmoset monkeys. *Curr Biol* 27: 1844–1852.
- Tanaka T, Sugiura H, Masataka N, 2006. Cross-sectional and longitudinal studies of the development of group differences in acoustic features of coo calls in two groups of Japanese macaques. *Ethology* 112:7–21.
- Whitham JC, Gerald MS, Maestriperi D, 2007. Intended receivers and functional significance of grunt and girney vocalizations in free-ranging female rhesus macaques. *Ethology* 113:862–874.
- Wilson M, Wilson TP, 2005. An oscillator model of the timing of turn-taking. *Psychon B Rev* 12:957–968.
- Yoshida S, Okanoya K, 2005. Evolution of turn-taking: a bio-cognitive perspective. *Cogn Stud* 12:153–165.
- Zarco W, Merchant H, Prado L, Mendez JC, 2009. Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *J Neurophysiol* 102:3191–3202.
- Zimmermann U, Rheinlaender J, Robinson D, 1989. Cues for male phonotaxis in the duetting bushcricket *Leptophyes punctatissima*. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 164:621–628.

