Complex rules of vocal sequencing in marmoset monkeys

Junfeng Huang[1,2], He Ma[1], Yongkang Sun[1], Liangtang Chang[1], Neng Gong[1,3*]

[1]Institute of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China.

[2]Lingang Laboratory, Shanghai 200031, China.

[3]Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai 201210, China.

*Correspondence: ngong@ion.ac.cn (N.G.)

Lead Contact: Neng Gong (ngong@ion.ac.cn)

ABSTRACT

Vocal sequencing is a key element in human speech. Songbirds have been widely studied as an animal model to investigate neural mechanisms of vocal sequencing, due to the complex syntax of syllable sequences in their songs. However, songbirds are phylogenetically distant from humans. So far, there is little evidence of complex syntactic vocalizations in non-human primates. Here, we analyze phee sounds produced by 160 marmoset monkeys either in isolation or during vocal turn-taking and reveal complex sequencing rules at multiple levels. First, phee syllables exhibited consistent interval patterns among different marmosets, allowing categorization of calls with single and closely spaced 2-4 syllables into 4 grades. Second, the ordering of sequential calls followed distinct probabilistic rules that preferring repetition of the same-grade call and then transition between calls of adjacent grades, but not skip-grade transition. Moreover, inter-call intervals depended on the transition direction. Third, specific AB^nA call patterns were discovered to be prominent in long call sequences, and their occurrence exhibited a power-law decrease with increasing "n", reflecting a long-range sequencing rule in the dependence of later calls on the pattern of earlier calls. Finally, syllable and call intervals as well as call compositions were significantly modified during vocal turn-taking. This complex syntax of vocal sequences in marmosets offers opportunities for understanding the evolutionary origin and neural mechanisms of grammatical complexity in human language.

KEYWORDS

Vocal sequencing; marmoset; vocal turn-taking; call transition; non-human primate; language

INTRODUCTION

Human speech is composed of ordered sequences of syllables, which convey meaning by forming words via their distinct combinations. Monosyllabic and multisyllabic words can be further integrated into sentences according to complex grammatical rules. No other mammalian species produce vocalizations that are as richly structured as human language. Many songbird species could produce songs that have complex syntax with syllable sequences following certain rules [1-4], which are flexible in different contexts [5]. Therefore, songbirds have been widely studied as an animal model to investigate the evolution and neural mechanisms of complex vocal sequencing [6, 7]. Since songbirds are phylogenetically distant from humans, it would be of interest to examine whether complex syntactic vocalization also exists in non-human primates (NHPs).

Common marmoset (*Callithrix jacchus*), a New World monkey species, has attracted much attention as a useful model for studying social behaviors, especially in vocal communication and monogamous cooperative breeding [8-10]. Marmoset monkeys have a complex vocal repertoire consisting of rich call types both in captivity and natural environment [11, 12]. They show typical vocal turn-taking, and produce trill and phee calls for close and long-distance communication, respectively [13, 14]. Notably, calls made by infant marmosets in the family group undergo dramatic changes during the first two months after birth, in a manner that depends on both physical maturation and contingent parental feedback [15, 16]. The influence of parental feedback on vocal development was also found in a juvenile stage (3 postnatal months) [17, 18]. These findings indicate that marmoset represents a good model for studying early vocal development in primates. In this study, we further demonstrated that marmoset vocalization could be used for studying complex syntactic vocal sequencing in non-human primates.

RESULTS

Interval Patterns at the Syllable Level

When kept in isolation, marmosets produce phee calls, which are known to represent long-distance contact calls for other conspecifics [14]. In this study, we recorded vocalizations of 110 adult marmosets (54 males and 56 females; age 6.1 ± 0.3 years; weight 390 ± 7 g, SEM) in isolation. Each

marmoset was recorded for 30 - 90 min (see Methods, Figure1A), and nearly all calls recorded were phee calls. In total, we acquired 28,622 discrete phee syllables (Figure 1B). We measured inter-syllable intervals (ISIs) between all adjacent syllables (n = 28,512 events) and found two distinct groups of ISIs by plotting all ISIs in a logarithmic scale (Figure 1C). The ISI distribution displayed a bimodal pattern fitting well with two Gaussian distributions, with two distinct peaks of occurrence percentages at 0.32 s and 10.67 s, respectively (Figure 1D). This ISI pattern of phee syllables in marmosets was consistent with a previous finding [19]. We then examined the ISI distribution of phee syllables for each marmoset and found that such bimodal pattern was largely invariant among marmosets, as shown by the heat plot of ISI distributions for all 110 animals (Figure 1E). Examples of ISI distributions for three marmosets (#27, #57, #75; Figures 1F - 1H) showed that each marmoset exhibited two peak values of ISIs that were slightly variable. These peak values for all 110 marmosets were clearly separated into two groups with the median at 0.35 s and 10 s (Figure 1I), close to that found for all phee syllables plotted together as one group (Figure 1D). Thus, phee syllables have consistent interval patterns in all marmosets.

Ordinal rules at the call level

According to the overall ISI distribution, the interval of 1 s (log value = 0) could well separate ISIs into two groups of short and long ISIs (Figure 2A). We thus define closely spaced syllables with the median short interval of ~0.35 s to be a call (or a word), and long intervals with a median of ~10 s represent intervals between calls. All 28,622 phee syllables recorded were grouped into 12,863 phee calls, which were graded by the number of syllables contained in the call, with N-phee representing phee call containing N syllables. We found that 2-phee was the most common (57.9%), followed by 3-phee (25.9%), 1-phee (12.1%), and 4-phee (3.7%) (Figure 2B). Calls containing more than 4 syllables were rare (~0.4%, Figure 2B), thus were not included in the following analyses. Consistent with the Zipf's law that was used to describe the complexity of vocal sequence [20-23], the proportion of occurrence for different ranks of N-phee calls showed a linear relationship in log-log plot, with a Zipf value of -1.86 (Figure 2C).

Through chunking of original phee syllable sequences, we obtained new phee call sequences, as shown by representative data from three marmosets in Figure 2D (See Figure S1 for data from all 110

marmosets). To examine the potential ordinal rule in phee call sequences, we used matrix to describe the frequency of call-grade transition between two sequential calls, for call sequences from all 110 marmosets (Figure 2E), and probabilities of transition into the same and different call grades are shown in the transition diagram in Figure 2F. We found that the occurrence percentage of transition into a call of the same grade ("Rep": repetition, 66.3%) was two-fold of that of transition to another grade ("Trans", 33.7%). Most (92%) of the latter was one-grade transition (Trans-One) that occurred between calls of adjacent grades, and rest was skip-grade transition (Trans-Skip, 8%, Figure 2F). Furthermore, occurrences of Trans-Up (transition to a higher grade) and Trans-Down (transition to a lower grade) were largely symmetrical, as shown by percentages of T_{AB} vs. T_{BA} in Figure 2G. By analyzing the inter-call intervals for all repetitions and one-grade transitions, we found that the interval to the next call showed a similar transition-direction dependence for all call grades (Figure 2H). Overall, the inter-call interval for repetition (11.2 s, median; n = 8408) was longer than that of Trans-Up (9.0 s, median; n = 2037, $P < 0.001$), and shorter than that of Trans-Down (16.3 s, median; n = 1887, $P < 0.001$) (Figure 2I). Thus, the ordering of marmoset phee calls follows distinct probabilistic rules that prefer repetition and one-grade transition, with inter-call intervals depending on the direction of transition.

Long-range rule and transition patterns in phee call sequences

The existence of a long-range sequencing rule for call transitions was further studied by examining the influence of the particular type of call transition on the occurrence of subsequent transitions. We first analyzed the occurrence frequency of various patterns of two sequential transitions, as shown by the frequency matrix in Figure 3A. We found that sequential repetitions ("Rep" to "Rep") occurred at the highest frequency, and Trans-Up or Trans-Down tended to be followed by "Rep" or a "Trans" of the opposite direction (Figure 3A). The long-range relationship between Trans-Up and Trans-Down transitions in a call sequence was further revealed by omitting the "Rep" between two "Trans" transitions, via grouping of sequential repeated calls into a single call (Figure 3B). The resulting frequency matrix containing only Trans-Up and Trans-Down (Figure 3B) showed that two sequential "Trans" transitions were much more likely to be opposite in direction. For example, when only "Trans" were counted, the overall average probability of 2-phee transiting to 1-phee and 3-phee were 0.41 and

0.59, respectively (Figure 3C). However, by examining the pattern of sequential two "Trans" transitions, we found that, when repeated 2-phee (2^n) was preceded by 1-phee (transition from 1-phee to 2-phee, T_{12}), the probability of transition to 1-phee (T_{21}, 0.71) was much higher than the average, whereas that of transition to 3-phee (T_{23}) dropped to 0.29 (Figure 3C). Similarly, prior 3-phee (T_{32}) led to an increased probability (140% of the average) of T_{23} and a reduced probability (41% of the average) of T_{21} (Figure 3C). The same pattern was also observed for transition from 3-phee to call of other grades (Figure 3C). Taken together, we found two typical transition patterns, AB^nA and AB^nC, in a long phee call sequence.

The AB^nA pattern consisted of two types, "Down-Up" (21^n2, 32^n3 and 43^n4) and "Up-Down" (12^n1, 23^n2 and 34^n3) (Figure 3D), whereas the AB^nC pattern consisted of "Down-Down" (43^n2 and 32^n1) and "Up-Up" (12^n3 and 23^n4) types (Figure 3D). The occurrence frequency of AB^nA (overall, 86%; individual average 84.7 $\pm$ 1.0 %, SEM) was significantly higher than that of AB^nC (overall, 14%; individual average 15.3 $\pm$ 1.0 %, SEM, n = 110, $P < 0.001$, Wilcoxon signed rank test) (Figure 3D). These results demonstrate the existence of long-range rule in marmoset vocal sequences. We then analyzed the number of B repetitions ("n") and found the time interval between the first (A) and last call (A or C) in AB^nA and AB^nC patterns increased linearly with the "n" value (range 1-10) in both patterns (Figure 3E). Moreover, the distribution of AB^nA and AB^nC patterns showed that most AB^nA and AB^nC occurred with relatively short A-A and A-C time intervals (26 s and 19 s at the peak frequency, respectively), while such tendency was more obvious in the distribution of AB^nA than that of AB^nC (Figure 3F). Consistently, both the frequency of AB^nA and AB^nC patterns decreased with increasing "n" values in a manner best fit with the power-law distribution, but AB^nA declined faster than AB^nC (Figures 3G and 3H). These different decline rates could also be revealed by the gradual decrease of the AB^nA to AB^nC frequency ratio with increasing "n" value (Figure 3I), suggesting a stronger long-range interaction in AB^nA than that in AB^nC pattern. Thus, the long-range ordering in phee call sequences is call-pattern specific and exhibits a power-law reduction in the temporal range.

Rules of vocal sequencing and antiphonal calling during turn-taking

Following the studies of phee call sequencing in isolation, we further examined whether these rules of vocal sequencing were changed during vocal turn-taking. Using 25 adult marmosets couples in various

families (age 4.5 ± 0.2 years; weight 403 ± 8 g; SEM, n = 50), we recorded vocalizations of each marmoset in the couple during vocal turn-taking separated by a curtain for ~60 min (Figure 4A， see Methods). In the absence of visual contact, nearly all antiphonal calls recorded were phee calls. First, we analyzed sequencing rules of each marmoset's own vocalizations during turn-taking. A total of 9905 phee syllables was acquired, and syllable density during vocal turn-taking (3.3 ± 0.3 per min, SEM, n = 50) was significantly lower than that in isolation (4.3 ± 0.3 per min, SEM, n = 110, P = 0.019, Mann-Whitney U test). The inter-syllable intervals during vocal turn-taking also distributed into two distinct groups, as shown by the heat plot of ISI distribution (Figure S2), but intra-call syllable intervals (first peak) were shorter and inter-call intervals (second peak) were longer than that found in isolation (Figure 4B). By the same criterion of 1-s threshold, all phee syllables were grouped into 4836 phee calls (call sequences were shown in Figure S3), and the composition of calls was markedly modified by vocal turn-taking (Figure 4C). Specifically, the overall proportion of 1-phee was elevated from 12.1% to 28.4%, and the corresponding Zipf value was changed from -1.86 to -1.38 (Figure 4D). This modification of call composition was also observed in each individual, as shown by the averaged proportion of 1-phee significantly elevated from 13.1 ± 1.6 % (SEM, n = 110) to 23.0 ± 3.4% (SEM, n = 50, P = 0.017, Mann-Whitney U test). By analyzing the relationship between two sequential calls as well as transition patterns, we found that the call ordering, transition probabilities and long-range effects in AB^nA patterns were nearly identical to that observed in isolation (Figures S4 and S5). The shortening of intra-call syllable intervals was similar across different call grades (Figure 4E). The prolonged inter-call intervals still depended on the transition direction: Comparing to that for repetitions (15.5 s, median; n = 3116), the intervals were shorter for Trans-Up (12.5 s, median; n = 773, P< 0.001) and longer for Trans-Down (24.0 s, median; n = 722, P < 0.001) (Figures 4F and S6). These results indicate that sequencing rules of phee calls during vocal turn-taking are generally similar to those found in isolation, but the syllable and call intervals, as well as call compositions were significantly modified.

We then examined the rules for antiphonal calling during vocal turn-taking. Two sequential calls made by different monkeys was considered to be a call pair in communications. We found a total of 2135 call pairs and analyzed the distribution of reply latency in each pair (Figure 4G). We confirmed the presence of vocal turn-taking by showing that the distribution of latencies was distinctly different

from that found after random permutation of marmosets' vocal sequences in each pair (Figure 4G). According to the distribution of latencies, a call pair with a reply latency less than 12 s was considered as an effective antiphonal call pair, similar to that used previously [24]. A total of 1443 effective antiphonal call pairs were observed, and the frequency (Figure 4H) and occurrence percentage (Figure 4I) for each type of call pair showed that marmosets preferred to use 1-phee and 2-phee for antiphonal calling. Furthermore, the reply call tended to be in the same or adjacent but not non-adjacent grade to the initiating call (Figure 4I). The latency for each initiating call eliciting a reply call of any grade was found to be significantly different for 4 grades of initiating calls, with 1-phee eliciting the more rapid reply than other grades of initiating calls (Figures 4J and 4K, see original data in Figure S7). These results revealed the rule of antiphonal calling during vocal turn-taking in marmosets (Figure 4L), suggesting potential meanings carried by phee calls under different social contexts.

DISCUSSION

By recording of vocalizations from hundreds of marmosets in isolation and during vocal turn-taking, we performed detailed analyses of their vocal sequencing and uncovered sequencing rules at multiple levels (Figure 5). First, interval patterns of phee syllables are consistent among different marmosets, and closely spaced syllables could be grouped into calls ("chunking"). Second, for two sequential calls, the probability of transition is highest for repetition of the same grade, followed by one-grade transition, and lowest for skip-grade transition. Third, for long-range sequencing rules, AB^nA patterns are much more prominent than AB^nC patterns, reflecting the dependence of later calls on the pattern of earlier calls. Finally, syllable and call intervals, as well as call compositions are significantly modified during vocal turn-taking. These findings showed that marmoset vocal sequences follow complex rules. Sequencing complexity of vocalization has previously been shown mainly in bird songs [1-4]. Although marmoset vocalization rules we found are quite simple in comparison to the complex syntax of human languages, they point to potential existence of richly structured vocal communication among marmoset monkeys.

We found that sequencing rules during vocal turn-taking generally followed those observed in isolation, including the probabilistic rule of call transitions and the long-range call sequencing.

However, the syllable and call intervals as well as the call grade compositions were significantly modified during turn-taking. This suggests that vocal sequencing in marmosets is probably determined by both the intrinsic constraints in vocal production and external influences by social contexts. During vocal turn-taking, the syllables within each call exhibited shorter spacing. This resulted in a shortening of the overall length of phee calls, an advantageous adaptation for reducing the overlap of vocal sounds during turn-taking (see also [25]). The latency of reply depended on the grade of initiating calls, and 1-phee elicited the fastest reply than higher grades of phee calls. This is in line with the notion that shorter latency in turn-taking facilitates marmoset vocal communication [13, 24]. We speculate that the higher usage of 1-phee during vocal turn-taking reflect the marmoset's willingness to actively communicate with each other, whereas phee calls containing different grades of syllables may convey messages under different social contexts. Skip-grade transitions are rare in marmosets' phee call sequences in isolation and during vocal turn-taking, a phenomenon that could be attributed to the physical constraint in vocal sound production. On the other hand, we noted that relative to grade of initiating call, marmosets rarely chose skip-grade reply call. This avoidance of both skip-grade transition and skip-grade reply may reflect voluntary control of vocalization at a higher cognitive level.

Complex rules of vocal sequencing at different levels reflect distinct brain mechanisms for sequence coding [26]. For example, the premotor nucleus HVC of songbirds is a key site for encoding probabilistic rules and long-range rules in the song syntax [2, 4]. Previous studies have shown that marmoset’s rich vocal repertoire has been valuable for studying neuronal mechanisms of auditory perception and vocal production [27-30]. For examples, marmosets could precisely control vocal behavior by rapidly interrupting ongoing vocalizations [31]. Their ability of motor planning for vocal production was suggested by the finding that the first syllable of the phee calls determines the call grade [32]. The ordinal rules for phee call sequences shown in this study indicate that the production of different grades of phee calls is also greatly influenced by previous calls. The vocal production in marmosets is controlled by auditory feedback [33, 34], and such feedback might account for the change of syllable and call intervals during vocal turn-taking shown in this study. Furthermore, vocal and locomotor coordination in marmosets is related to changes in the autonomic nervous system [35]. “Mayer wave”, an 0.1 Hz oscillation of the autonomic nervous system may be responsible for the overall 10-s inter-call intervals in marmosets [36]. Other higher-level brain mechanisms must underlie

the transition-direction dependence of intervals between sequential calls, as well as different syllable and call intervals in isolation vs. vocal turn-taking conditions. Most previous studies on marmoset vocalization mainly focused on mechanisms of sound production and neural processing at auditory and motor cortices. The neural mechanism for cognitive control of complex vocal sequencing in primates remains largely unknown.

In summary, our findings provide strong evidence for the existence of complex syntactic vocalization in a non-human primate and enrich the previous concept of primate vocal pattern generation [37]. Accompanying the evolution of the vocal production apparatus [38], there is an emergence of voluntary cognitive control mechanisms for complex vocal sequencing in primates [39, 40]. Our study underscores the usefulness of marmoset monkeys for studying cognitive control of vocal sequencing that may shed new light into the neural basis of grammatical complexity in human language.

ACKNOWLEDGEMENTS

We thank Mu-ming Poo, Liping Wang and Zhen-Hua Ling for helpful discussions and insightful comments on this manuscript, Yishan Xie, Ruixin An, Yuan Xin, Xiangrui Li and the marmoset facility at CAS Center for Excellence in Brain Science and Intelligence Technology for assistance in sound recordings and analyzing. This work was supported by Science and Technology Innovation 2030-"Brain Science and Brain-inspired Research" Major Project Grant No. 2021ZD0203900, Shanghai Municipal Science and Technology Grant No. 22ZR1481500, NSFC Project 31871068, "Strategic Priority Research Program" of the Chinese Academy of Sciences, Grant No. XDB32010000, Shanghai Municipal Science and Technology Major Project, Grant No. 2018SHZDZX05, and CAS Key Technology Talent Program to N.G.

AUTHOR CONTRIBUTIONS

N.G. conceived and designed the study. J.H., H.M., Y.S. and L.C. performed sound recordings. J.H. and H.M. analyzed data and prepared the figures. N.G. and J.H. wrote the manuscript. All authors have read and approved the manuscript submission.

DECLARATION OF INTERESTS

The authors declare no competing interests.

METHODS

Subjects

In this study, we recorded vocalizations of 110 adult common marmosets (54 males and 56 females; aged 1.5 - 13.5 years and weighted 220 - 625 g) in isolated condition and 25 marmoset couples (25 males and 25 females; aged 1.9 - 9.5 years and weighted 308 - 516 g; 7 of them were also recorded in isolation) in paired condition. All marmosets were captive and lived in family groups. Each marmoset family was housed in a wire-mesh cage (L × W × H: 0.90 m × 0.80 m × 0.85 m), equipped with a sleeping box and other enrichment materials. Marmosets had ad libitum access to food and water, and breeding rooms were maintained at a temperature range from 26℃ to 30℃ and a 12-h: 12-h light-dark cycle. Animal care and experimental procedures were approved by the Animal Care and Use Committee at the Institute of Neuroscience, Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences.

Experimental setup and vocal recording

All vocal recordings were conducted in a soundproof recording room (L × W × H: 4.5 m × 1.5 m × 2.2 m) from 2013 to 2022. For the recording in the isolated condition, each marmoset was transported by a transfer box from the breeding room to the recording room, and then placed in a wire-mesh cage (L × W × H: 0.90 m × 0.80 m × 0.85 m) at the height of 1.20 m above the floor. A directional condenser microphone (Audio-Technica AT2031) was placed directly in front of the cage at a distance of 0.35 m, and the recorded audio signal was digitized through an audio interface (Icon Utrack or Roland Octa-Capture UA-1010) at 48 kHz sample rate with 16 bits by Adobe Audition software. Each marmoset was recorded for 30 - 90 min with the mean time of 61.6 min. In the paired condition, two marmosets in a couple were removed to the recording room and placed separately in two cages about 3.0 m apart, with a microphone placed directly 0.35 m in front of each cage. An opaque black curtain was hung between the cages to occlude visual contact. Each couple was recorded for ~60 min.

Vocalization detection and classification

After a high-pass filtering of audio signals at 3 kHz, a custom-made MATLAB script was used to detect the beginning and ending of each syllable by applying an energy threshold. Then, the types of vocalizations were annotated manually in Praat software based on their spectrograms. The detection and classification results were cross-checked by 2 - 3 researchers. In both isolated and paired conditions, recorded vocalizations were almost phee calls, the specific long-distance contact call in marmosets. In total, we had recorded 28,622 and 9,905 phee syllables in the isolated and paired condition, respectively.

Vocalization analysis

All data analyses of vocalizations were performed in MATLAB by custom-made scripts.

Distribution of Inter-syllable interval. Inter-syllable interval (ISI) was defined as the silence period between two sequential syllables. After plotting the ISIs in a logarithmic (log 10) scale, the ISI distribution displayed a bimodal pattern with two peaks, each mode of which was fitted with one Gaussian function:

$$f(x) = a * \mathrm{e}^{\left(-\frac{(x-\mu)^2}{2b^2}\right)}$$

For each individual marmoset, the ISI distribution was smoothed by averaging the occurrence percentage of the nearby three bins (Figures. 1F - 1H), and two peak ISI values in the distribution were measured. All these two peak values for 110 marmosets were plotted as two groups and their distributions were smoothed with a normal kernel function in Figure. 1I.

Zipf value. Zipf value was used to describe the occurrence of N-phee calls. In linguistics, the Zipf's law states that the occurrence proportion of different words is inversely proportional to its rank of the proportion with the slope (also called Zipf value) about -1.00 in the log-log scale for human languages.

$$log_{10}\, Proportion_i = a * log_{10}\, Rank_i + b$$

where $Proportion_i$ and $Rank_i$ represents respectively the proportion of occurrence and its rank for the N-phee call containing "i" syllables. The coefficient a is the Zipf value.

Distribution of AB^{n}A and AB^{n}C patterns. The distribution of the occurrence frequency of AB^{n}A or AB^{n}C pattern at various "n" values was fitted with a power-law function:

$$f(x) = ax^b$$

where the value of exponent "b" indicates the decline rate.

Time window of antiphonal calling. During vocal turn-taking, two sequential phee calls made by different monkeys were considered to be a call pair. The interval time between the ending of the initiating call and the beginning of the reply call was regarded as reply latency, with negative value representing overlapping of two calls. We permutated the reply sequence by shuffling the order of inter-call intervals randomly for 1,000 times in one marmoset's phee call sequence of each couple, and then compared the distribution of latency of all 2,135 call pairs to that of the permuted data set. The time window of antiphonal calling was determined as 0 - 12 s, in which the occurrence of call pair in experimental data was more than that of permuted data set.

Statistical analysis

All statistical analyses were accomplished with SPSS Version 21.0. Linear mixed-effects model (LMM) was conducted to compare intra-call syllable intervals and inter-call intervals among groups. The syllable and call intervals were log-transformed (log10) to meet the normality assumption of LMM and Bonferroni corrections were applied for multiple comparisons. Kolmogorov-Smirnov test was conducted to compare different cumulative distributions. Mann-Whitney U test was conducted to compare the syllable density or the proportion of calls. Wilcoxon signed rank test was conducted to compare the occurrence frequency of transition patterns. All statistical tests were two-tailed.

SUPPLEMENTAL INFORMATION

Figures S1 - S7

REFERENCES

1. Okanoya, K. (2004). The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. Annals of the New York Academy of Sciences *1016*, 724-735.
2. Zhang, Y.S., Wittenbach, J.D., Jin, D.Z., and Kozhevnikov, A.A. (2017). Temperature manipulation in songbird brain implicates the premotor nucleus HVC in birdsong syntax. Journal of Neuroscience *37*, 2600-2611.
3. Markowitz, J.E., Ivie, E., Kligler, L., and Gardner, T.J. (2013). Long-range order in canary song. PLoS computational biology *9*, e1003052.
4. Cohen, Y., Shen, J., Semu, D., Leman, D.P., Liberti, W.A., Perkins, L.N., Liberti, D.C., Kotton, D.N., and Gardner, T.J. (2020). Hidden neural states underlie canary song syntax. Nature *582*, 539-544.
5. Veit, L., Tian, L.Y., Hernandez, C.J.M., and Brainard, M.S. (2021). Songbirds can learn flexible contextual control over syllable sequencing. Elife *10*, e61610.
6. Doupe, A.J., and Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. Annual review of neuroscience *22*, 567-631.
7. Lipkind, D., Marcus, G.F., Bemis, D.K., Sasahara, K., Jacoby, N., Takahasi, M., Suzuki, K., Feher, O., Ravbar, P., and Okanoya, K. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. Nature *498*, 104-108.
8. Miller, C.T., Freiwald, W.A., Leopold, D.A., Mitchell, J.F., Silva, A.C., and Wang, X. (2016). Marmosets: a neuroscientific model of human social behavior. Neuron *90*, 219-233.
9. Saito, A. (2015). The marmoset as a model for the study of primate parental behavior. Neuroscience Research *93*, 99-109.
10. Huang, J., Cheng, X., Zhang, S., Chang, L., Li, X., Liang, Z., and Gong, N. (2020). Having infants in the family group promotes altruistic behavior of marmoset monkeys. Current Biology *30*, 4047-4055. e4043.
11. Agamaite, J.A., Chang, C.-J., Osmanski, M.S., and Wang, X. (2015). A quantitative acoustic analysis of the vocal repertoire of the common marmoset (Callithrix jacchus). The Journal of the Acoustical Society of America *138*, 2906-2928.

12. Snowdon, C.T. (1989). Vocal communication in New World monkeys. Journal of Human Evolution *18*, 611-633.
13. Miller, C.T., and Wang, X. (2006). Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. Journal of comparative physiology A *192*, 27-38.
14. Liao, D.A., Zhang, Y.S., Cai, L.X., and Ghazanfar, A.A. (2018). Internal states and extrinsic factors both determine monkey vocal production. Proceedings of the National Academy of Sciences *115*, 3978-3983.
15. Takahashi, D.Y., Fenley, A.R., Teramoto, Y., Narayanan, D., Borjon, J.I., Holmes, P., and Ghazanfar, A.A. (2015). The developmental dynamics of marmoset monkey vocal production. Science *349*, 734-738.
16. Pistorio, A.L., Vintch, B., and Wang, X. (2006). Acoustic analysis of vocal development in a New World primate, the common marmoset (Callithrix jacchus). The Journal of the Acoustical Society of America *120*, 1655-1670.
17. Gultekin, Y.B., and Hage, S.R. (2017). Limiting parental feedback disrupts vocal development in marmoset monkeys. Nature communications *8*, 1-9.
18. Gultekin, Y.B., and Hage, S.R. (2018). Limiting parental interaction during vocal development affects acoustic call structure in marmoset monkeys. Science Advances *4*, eaar4012.
19. Takahashi, D.Y., Narayanan, D.Z., and Ghazanfar, A.A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. Current Biology *23*, 2162-2168.
20. Zipf, G.K. (1949). Human behavior and the principle of least effort: an introd. to human ecology.
21. Cancho, R.F.I., and Solé, R.V. (2003). Least effort and the origins of scaling in human language. Proceedings of the National Academy of Sciences *100*, 788-791.
22. McCOWAN, B., Hanser, S.F., and Doyle, L.R. (1999). Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. Animal behaviour *57*, 409-419.
23. Gultekin, Y.B., Hildebrand, D.G., Hammerschmidt, K., and Hage, S.R. (2021). High plasticity in marmoset monkey vocal development from infancy to adulthood. Science Advances *7*, eabf2938.
24. Miller, C.T., Beck, K., Meade, B., and Wang, X. (2009). Antiphonal call timing in marmosets is behaviorally significant: interactive playback experiments. Journal of Comparative Physiology A *195*, 783-789.

25. Yamaguchi, C., Izumi, A., and Nakamura, K. (2010). Time course of vocal modulation during isolation in common marmosets (Callithrix jacchus). American Journal of Primatology *72*, 681-688.
26. Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. Neuron *88*, 2-19.
27. Eliades, S.J., and Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature *453*, 1102-1106.
28. Eliades, S.J., and Wang, X. (2019). Corollary discharge mechanisms during vocal production in marmoset monkeys. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging *4*, 805-812.
29. Zeng, H.-h., Huang, J.-f., Li, J.-r., Shen, Z., Gong, N., Wen, Y.-q., Wang, L., and Poo, M.-m. (2021). Distinct neuron populations for simple and compound calls in the primary auditory cortex of awake marmosets. National Science Review *8*, nwab126.
30. Cerkevich, C.M., Rathelot, J.-A., and Strick, P.L. (2022). Cortical basis for skilled vocalization. Proceedings of the National Academy of Sciences *119*, e2122345119.
31. Pomberger, T., Risueno-Segovia, C., Löschner, J., and Hage, S.R. (2018). Precise motor control enables rapid flexibility in vocal behavior of marmoset monkeys. Current Biology *28*, 788-794. e783.
32. Miller, C.T., Eliades, S.J., and Wang, X. (2009). Motor planning for vocal production in common marmosets. Animal Behaviour *78*, 1195-1203.
33. Eliades, S.J., and Tsunada, J. (2018). Auditory cortical activity drives feedback-dependent vocal control in marmosets. Nature communications *9*, 1-13.
34. Hage, S.R. (2020). The role of auditory feedback on vocal pattern generation in marmoset monkeys. Current Opinion in Neurobiology *60*, 92-98.
35. Gustison, M.L., Borjon, J.I., Takahashi, D.Y., and Ghazanfar, A.A. (2019). Vocal and locomotor coordination develops in association with the autonomic nervous system. Elife *8*, e41853.
36. Zhang, Y.S., and Ghazanfar, A.A. (2020). A hierarchy of autonomous systems for vocal production. Trends in neurosciences *43*, 115-126.

37. Fischer, J., and Hage, S. (2019). Primate vocalization as a model for human speech: scopes and limits. Human language: from genes and brains to behavior, 639-656.

38. Ghazanfar, A.A., and Rendall, D. (2008). Evolution of human vocal production. Current biology *18*, R457-R460.

39. Jürgens, U. (2002). Neural pathways underlying vocal control. Neuroscience & Biobehavioral Reviews *26*, 235-258.

40. Hage, S.R., and Nieder, A. (2016). Dual neural network model for the evolution of speech and language. Trends in neurosciences *39*, 813-829.

FIGURE 1

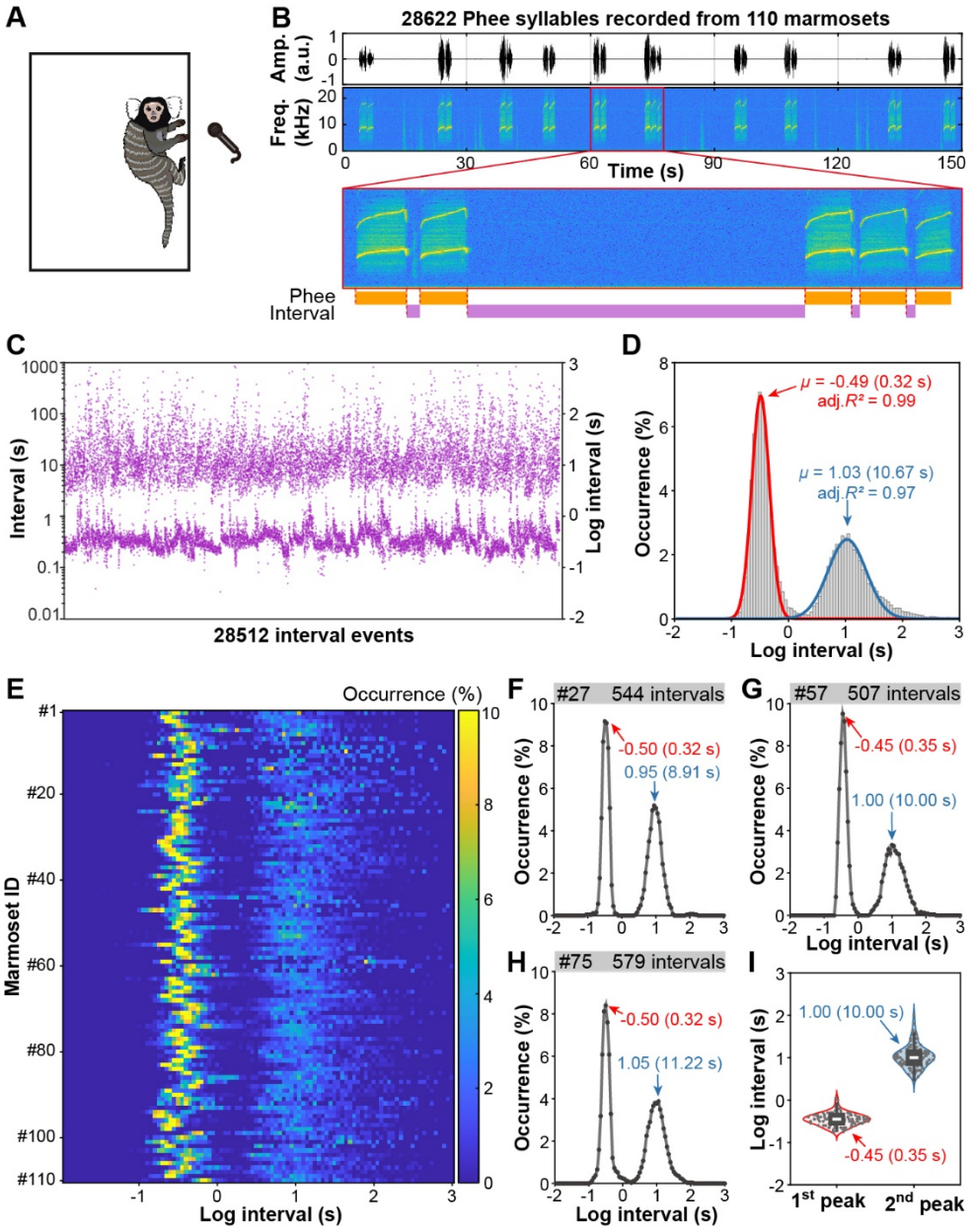

Figure 1. Consistent interval patterns of phee syllables in marmoset vocalizations.
(A) An illustration of the call recording method in which a marmoset was isolated in a cage in the sound-proof room and recorded for 30 - 90 minutes. (B) A typical example showing the recorded phee syllable sequence and the measurement of inter-syllable intervals (ISIs). (C) Plotting of all data points of ISIs in a log scale. (D) Distribution of ISIs in log scale and the fitting curves of Gaussian functions. (E) The heat plot of ISI distributions for all 110 animals. (F to H) Examples of ISI distributions for three marmosets #27, #57, and #75. (I) Plotting of two peak values in the ISI distribution for all 110 marmosets. White line represents the median.

FIGURE 2

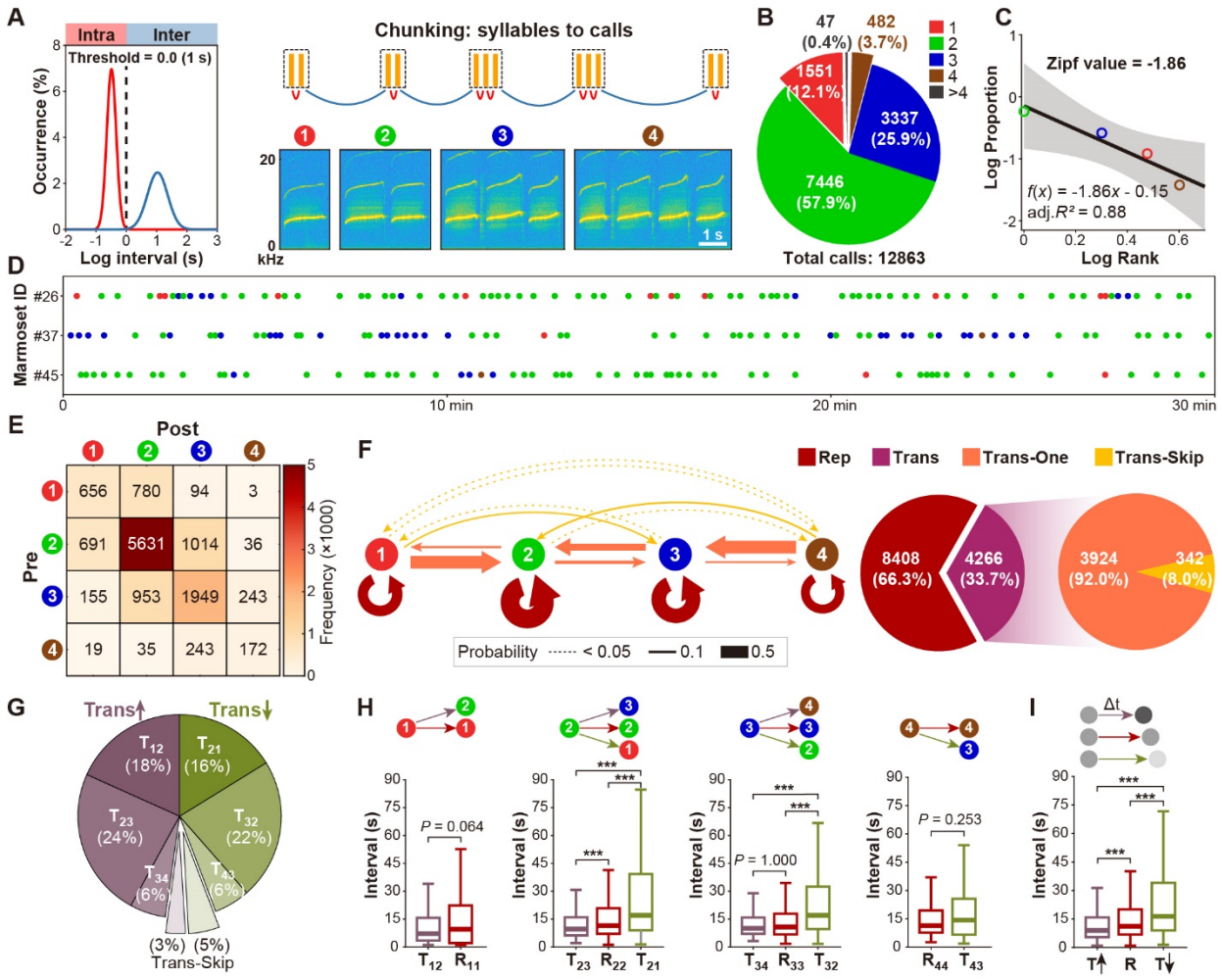

Figure 2. Ordering of sequential phee calls followed distinct probabilistic rules.

(A) According to a 1 s threshold shown in the fitting curve of ISI distribution, closely spaced phee syllables could be grouped together into different phee calls. (B) Pie diagram showing the proportion of each grade of phee call. (C) Distribution from more to less of different call grades in log-log scale. Zipf value was calculated. Shade area: 95% confidence interval. (D) Three examples showing the new phee call sequences through chunking of phee syllable sequences obtained from Marmoset #26, #37, and #45. (E) Matrix showing the frequency of all transitions between two sequential calls including data from all 110 marmosets. (F) Transition diagrams for call sequences. Arrow line color and thickness represents the type and the probability of transition, respectively. Pie diagram showing the occurrence percentage of each transition type. "Rep": repetition; "Trans": transition to a call of different grade; Trans-One: one-grade "Trans"; Trans-Skip: skip-grade "Trans". (G) Pie diagram showing the occurrence percentage of each Trans-Up (transition to a call of higher grade, left) or Trans-Down (transition to a call of lower grade, right). T_{AB}: transition from call A to B. (H) Box-plot diagrams showing medians of inter-call intervals for all specific transition types. (I) Overall data showing that inter-call intervals were shorter in Trans-Up, longer in Trans-Down than that of repetitions (Median; ***$P < 0.001$, linear mixed-effects model). T: "Trans"; R: "Rep".

See also Figure S1.

FIGURE 3

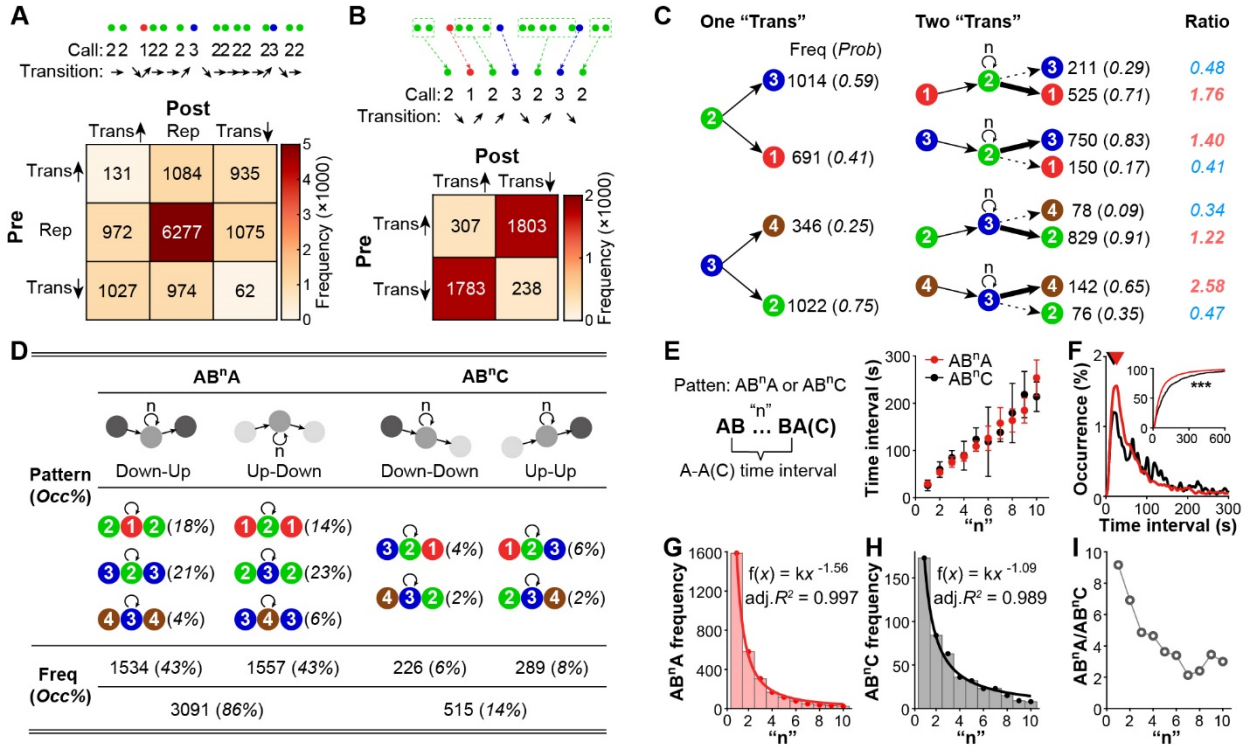

Figure 3. Long-range rule and transition patterns in phee call sequences.

(A) Transition direction information ("Rep", Trans-Up and Trans-Down) was extracted from call sequences, and the relationship between transition types was shown in frequency matrix. (B) Sequential repeated calls were grouped into a single call. Frequency matrix containing only Trans-Up and Trans-Down after omitting "Rep". (C) Frequencies and probabilities of transiting from 2-phee or 3-phee to the next call when only one "Trans" was considered and when sequential two "Trans" were considered showing the effect of the previous "Trans" on the next one, represented by ratio of the two probabilities. (D) The frequency and occurrence percentage of each transition pattern (Down-Up, Up-Down, Down-Down, Up-Up) showing a major AB^{n}A pattern in phee call sequences. (E) The A-A and A-C time intervals in AB^{n}A and AB^{n}C patterns increased linearly with the "n" value (Median ± SEM). (F) Distributions of the frequency of AB^{n}A and AB^{n}C patterns at various time intervals. Arrow represents the value of peak frequency. Cumulative distributions are shown in the inset (***$P < 0.001$, Kolmogorov-Smirnov test). (G and H) Distributions of the frequency of AB^{n}A and AB^{n}C patterns at various "n" values and the fitting curves of power-law function. (I) The ratio of the occurrence frequency of AB^{n}A to AB^{n}C at various "n" values.

FIGURE 4

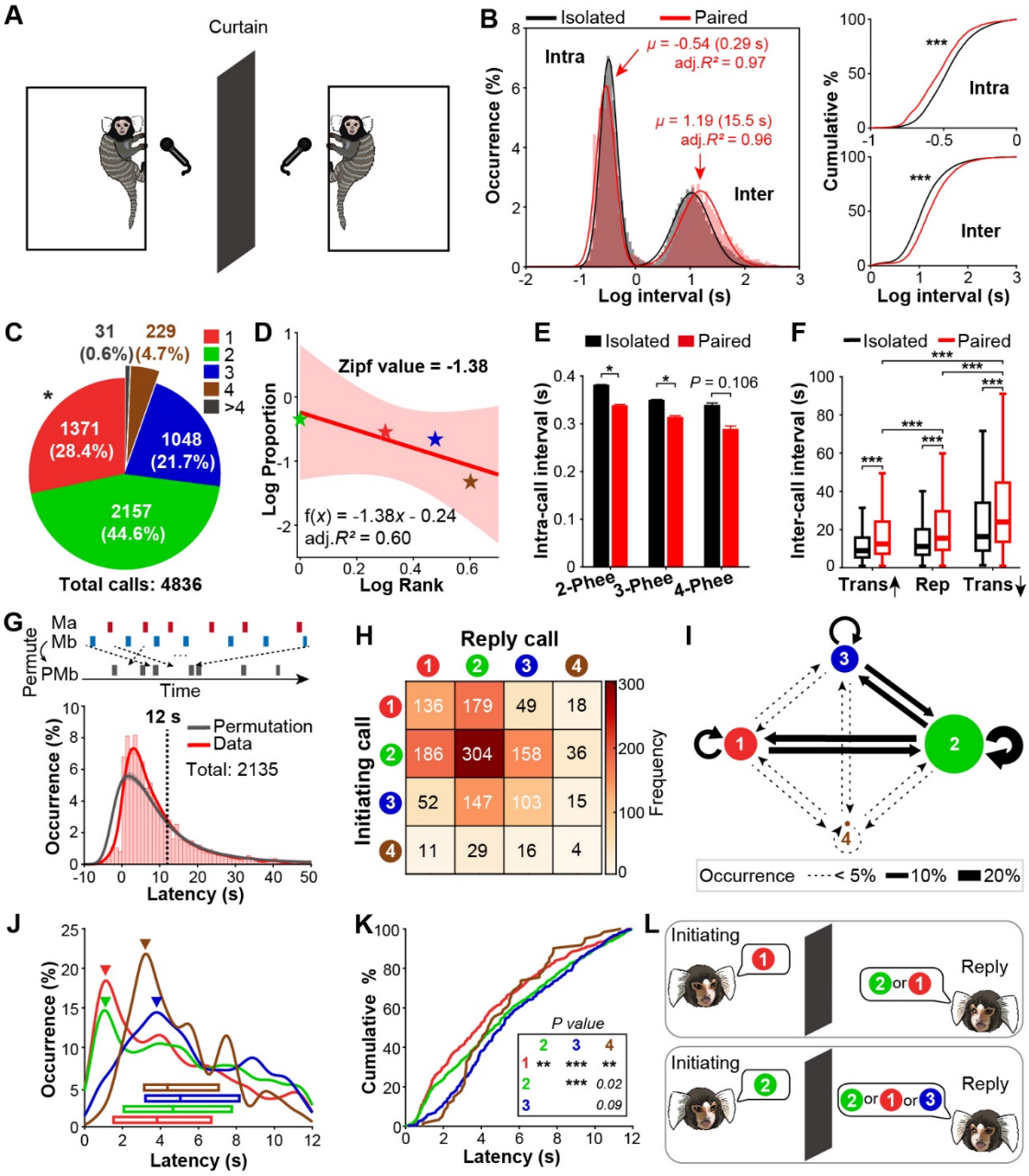

Figure 4. Rules of vocal sequencing and antiphonal calling during turn-taking.

(A) An illustration of the call recording method during vocal turn-taking between two marmosets. (B) Distributions of inter-syllable intervals in log scale and Gaussian fitting curves in both isolation and vocal turn-taking conditions. Right panel: cumulative distributions of intra-call syllable intervals and inter-call intervals (***$P < 0.001$, Kolmogorov-Smirnov test). (C) Pie diagram showing the proportion of each grade of phee call during turn-taking (*$P < 0.05$, Mann-Whitney U test). (D) Distribution from more to less of different call grades in log-log scale. Zipf value was calculated. Shade area: 95% confidence interval. (E) For each grade of phee call, intra-call syllable intervals were shorter during turn-taking than those in isolation (Mean ± SEM; *$P < 0.05$, linear mixed-effects model). (F) Inter-call intervals of different transition types in both isolation and turn-taking conditions (Median; ***P

< 0.001, linear mixed-effects model). (G) Distribution of reply latencies between two sequential calls made by two monkeys for both original and randomly permutated sequences. A call pair with a latency less than 12 s (dash line) was considered as an effective antiphonal call pair. Gray shade area: 95% confidence interval. (H) Frequency matrix showing the relationship between two sequential calls made by different marmosets in all effective antiphonal call pairs. (I) Reply diagrams in all effective antiphonal call pairs. Dot size represents the proportion of initiating call with different grades. Arrow line thickness represents the occurrence percentage of each call pair. (J) Distributions and medians of the latencies for calls of various grades to elicit a reply. Arrow represents the value of peak occurrence percentage. (K) Cumulative distributions of the latencies for calls of various grades to elicit a reply (**P < 0.01, ***P < 0.001, Kolmogorov-Smirnov test). (L) An illustration of antiphonal calling between two marmosets showing that marmosets prefer to use 1-phee and 2-phee calls as initiating calls, as well as the composition of reply calls elicited by 1-phee and 2-phee calls.

See also Figures S2 - S7.

FIGURE 5

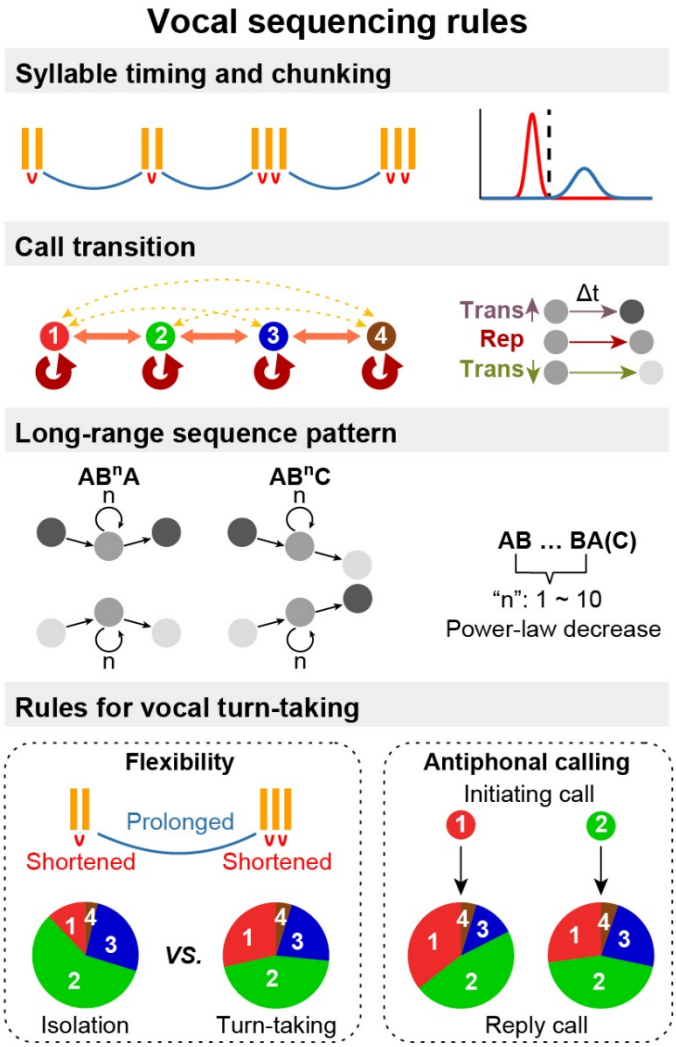

Figure 5. Summary of vocal sequencing rules found in marmoset phee vocalizations at multiple levels.

According to our study, marmosets produce phee calls following complex rules at four levels: (1) distinct interval patterns and "chunks" of syllables; (2) the transition between two sequential calls; (3) transition patterns occurred in long-range sequences; (4) vocal flexibility and antiphonal calling during turn-taking.