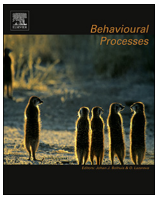
Behavioural Processes 162 (2019) 55–63

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc



Auditory sequence perception in common marmosets (*Callithrix jacchus*)

Masumi Wakita

Cognitive Neuroscience Section, Primate Research Institute, Kyoto University, Kanrin 41-2, Inuyama, Aichi, 484-8506, Japan

ARTICLE INFO

Keywords:
Auditory discrimination
Common marmoset
Monkey
Sequence
Temporal pattern

ABSTRACT

One of the essential linguistic and musical faculties of humans is the ability to recognize the structure of sound configurations and to extract words and melodies from continuous sound sequences. However, monkeys' ability to process the temporal structure of sounds is controversial. Here, to investigate whether monkeys can analyze the temporal structure of auditory patterns, two common marmosets were trained to discriminate auditory patterns in three experiments. In Experiment 1, the marmosets were able to discriminate trains of either 0.5- or 2-kHz tones repeated in either 50- or 200-ms intervals. However, the marmosets were not able to discriminate ABAB from AABB patterns consisting of A (0.5-kHz/50-ms pulse) and B (2-kHz/200-ms pulse) elements in Experiment 2, and A (0.5-kHz/50-ms pulse) and B (0.5-kHz/200-ms pulse) [or A (0.5-kHz/200-ms pulse) and B (2-kHz/200-ms pulse)] in Experiment 3. Consequently, the results indicated that the marmosets could not perceive tonal structures in terms of the temporal configuration of discrete sounds, whereas they could recognize the acoustic features of the stimuli. The present findings were supported by cognitive and brain studies that indicated a limited ability to process sound sequences. However, more studies are needed to confirm the ability of auditory sequence perception in common marmosets.

1. General introduction

The ability to recognize words from continuous speech is essential to understanding a spoken language. This process includes analyzing the order of the constituent sounds (e.g., syllables) and extracting meaningful units (e.g., words) from a sequence of speech sounds. In addition, the same syllables arranged in a different order can produce a new word with a different meaning. Music is also a rule-based system in which discrete elements (e.g., tones) are integrated into meaningful units (e.g., harmonious melodies). Different arrangements of the same tones result in different melodies. Thus, processing the temporal configuration of sounds is one of the fundamentals of humans' linguistic and musical faculties.

Prior studies have suggested that non-human primates can learn tonal sequences. Macaque monkeys can abstract the relative relationship between tones in auditory sequences when such contours are judged according to the ascending or descending direction of frequency change (Izumi, 2001; Brosch et al., 2004). Thus, macaques identify and memorize tones mostly according to their pitch contour, i.e., to the sequential up-and-down patterning between adjacent notes. However, Izumi (1999) has shown that a brief silence gap inserted between tones affects judgement of the pitch relation between two tones. In addition, Wright et al. (2000) have successfully trained two macaque monkeys to discriminate simple melodies. In addition, monkeys are able to generalize their responses to octave-transposed test melodies (cf. D'Amato and Salmon, 1984; D'Amato and Colombo, 1988). However, such generalization does not occur when the test melodies are arranged in different keys, even when the relative pitch relations between adjacent tones is preserved. This lack of generalization indicates that monkeys' discrimination of auditory sequences is largely controlled by local sensory cues, when available. Thus, temporal configurations of individual sounds may not be a primary cue to extract overall structural patterns in monkeys.

Over the past decade, however, many artificial grammar learning studies have indicated that monkeys can process auditory sequential rules (e.g., Fitch and Hauser, 2004; Newport et al., 2004; Saffran et al., 2008; Hauser and Glynn, 2009; Ravignani et al., 2013; Wilson et al., 2013; Wilson et al., 2015). These studies employed a habituation/dishabituation procedure: the monkeys were passively exposed to sequences of sounds structured according to a given rule and tested with acoustically new stimuli, which were arranged either congruently or incongruently with the original rule. Consequently, the monkeys showed a more frequent looking response toward the incongruent rather than the congruent patterns. Such results suggest not only that monkeys are sensitive to sound sequences but also that monkeys can learn a structural rule of sound sequences simply by being briefly exposed. Further, Wilson et al. (2015) have found that monkeys and humans have comparable brain regions in the ventral and inferior frontal

---

E-mail address: wakita.masumi.2e@kyoto-u.ac.jp.

https://doi.org/10.1016/j.beproc.2019.01.014
Received 21 September 2018; Received in revised form 26 December 2018; Accepted 31 January 2019

regions that play an important role in recognizing the differences between congruent and incongruent auditory patterns.

Recently, Beckers et al. (2017) and Ghirlanda (2017) have analyzed the experimental design in prior artificial grammar learning studies (Saffran et al., 2008; Wilson et al., 2013, 2015; Ravignani et al., 2013) and have found that the stimulus sequences contained biases in acoustic similarity between the habituation stimuli and test stimuli. Thus, the observed behavioral and brain responses may be explained in terms of matching of the auditory sensory features of the stimulus. In other words, the monkeys probably did not learn the structured auditory patterns on the basis of the configuration of individual elements during the habituation phase.

A habituation/dishabituation procedure is a convenient method to assess the cognitive ability to detect stimulus differences without explicit training. However, it is difficult to determine an association between monkeys' cognitive processes and their behavioral and brain responses, because the above mentioned artificial grammar learning studies did not require the monkeys to discriminate stimuli. In addition, successful discrimination between congruent and incongruent test sequence categories does not indicate that monkeys can discriminate between two different congruent (or two different incongruent) stimuli. Studies to date have collectively shown that monkeys can discriminate tonal sequences, but whether such discrimination is dependent on the perception of structured auditory patterns in terms of the temporal configuration of individual elements remains to be studied.

Therefore, it would be useful to explore the ability to process tonal sequences in monkeys under an experimentally controlled situation, in which the monkeys are trained to discriminate the sequences. European starlings have been found to discriminate auditory temporal patterns in an operant conditioning procedure (Gentner et al., 2006). Here, I conducted three experiments to test monkeys' ability to perceive structured auditory patterns on the basis of the temporal configuration of individual elements. Specifically, I trained two common marmosets to discriminate between two auditory patterns composed of common elements that were arranged differently. The common marmoset is a highly vocal monkey with a wide vocal repertoire (Pistorio et al., 2006; Bezerra and Souto, 2008), and some calls are combined. Therefore, this species can be considered an ideal non-human primate for studies on auditory perception.

2. Experiment 1: discrimination of sequences: one-frequency × one-interval

The principal aim of the current study was to determine whether common marmosets can perceive structured auditory patterns in terms of the temporal configuration of individual elements. To this end, in Experiment 2, the marmosets were trained to discriminate between ABAB and AABB patterns in which the A (a pure tone lasting 50 ms at 0.5 kHz) and B (a pure tone lasting 200 ms at 2 kHz) elements were regularly combined. However, if the marmosets were to confuse the A and B elements, the ABAB and AABB patterns would not be perceived as distinct sequences. Therefore, in Experiment 1, it was essential to confirm whether the marmosets were able to distinguish the acoustic features along the frequency and temporal dimensions of the elements that constituted the stimulus sequences in the following experiments.

3. Materials and methods

3.1. Subjects

Two experimentally naïve male common marmosets (*Callithrix jacchus*) participated in the current study. Both Cj190 and Cj195 were born in our laboratory, and each was housed with its own co-twin littermate. Cj190 was raised by veterinary staff. Both monkeys were 2 years of age and weighed 350 g at the start of the experiments. Marmosets were fed twice daily with conventional New World monkey pellets (SPS, Oriental Yeast, Tokyo, Japan) and vitamin supplements. They were additionally supplemented with various fruits, milk, and mealworms three times per week. Their daily consumption of water was not restricted. The training sessions were conducted between daily feedings.

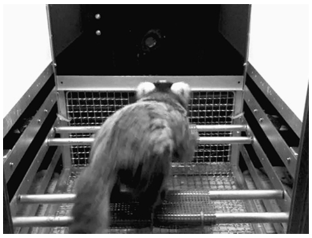


Fig. 1. Apparatus. A marmoset is shown moving from the stimulus perch to the response perch.

This study complied with the current laws of Japan. All experimental procedures and handling methods were performed in accordance with the "Guidelines for the Care and Use of Nonhuman Primates" of the Primate Research Institute, Kyoto University. The experiments were approved by the Animal Experiment Committee of Kyoto University.

3.2. Apparatus

The marmosets were trained in an experimental chamber [45 (h) × 30 (w) × 75 (d) cm] (Fig. 1). The chamber contained two perches—a stimulus perch and a response perch—which were coupled to photosensors positioned 5 cm above the floor to detect the location of the animal. The distance between the perches was 15 cm. The response perch was located 5 cm in front of the front panel. The front panel contained a round opening (3 cm in diameter) located 10 cm above the response perch, through which the marmosets were allowed to access food reward pellets (composed primarily of cornstarch and sugar). Normally, this window was closed with a solenoid shutter. Behind the front panel was a loudspeaker for stimuli presentation. The ceiling light could be turned on or off, and a water bottle was located on the rear panel. Water was freely available during training.

3.3. Stimulus

The discriminative stimuli were four different trains that consisted of pure tones (50-ms duration with 10-ms rise and fall times) with a frequency of either 0.5 or 2 kHz. The tones were repeated with intervals of either 50 or 200 ms of silence (Fig. 2A–D). The positive stimuli (S^+) were trains of pulses at 0.5 kHz presented with 50-ms silent intervals for Cj190 (Fig. 2A) and 2 kHz presented with 200-ms silent intervals for Cj195 (Fig. 2D). The intensities of the 0.5-kHz (65 dB SPL) and 2-kHz

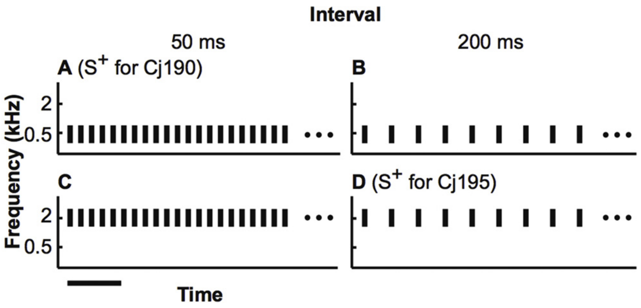


Fig. 2. Schematic representation of the stimuli in Experiment 1. Four types of stimuli were generated by repeating a pure tone of 50-ms duration at either 0.5- (A and B) or 2-kHz (C and D) frequencies with either 50- (A and C) or 200-ms (B and D) separations. Pulse trains illustrated in panels A and D were S^+ for Cj190 and Cj195, respectively. The stimulus trains were presented for up to 5 s. Scale bar = 500 ms.

(60 dB SPL) frequency tones were subjectively equal (Osmanski and Wang, 2011). Sound pressure was measured at the ear-level height of the marmosets in the position of the stimulus perch with a non-directional 1/2-inch condenser microphone (UC-53 A, Rion, Tokyo Japan) and a sound level meter (NA-42S, Rion, Tokyo Japan). During the training sessions, the marmosets' heads were not fixed. Although sensitivity to sound frequency is influenced by which direction the animals are facing, discriminability between 0.5 and 2 kHz tones was unlikely to be affected by the head direction because the frequency difference of these tones was large enough.

In Experiment 1, stimuli were defined by both frequency and temporal dimensions. By using such compound stimuli, the discriminability of those two distinct sensory features could be studied efficiently in one task. The temporal dimension of the stimulus trains was determined by using an empty interval, such that the temporal evaluation was not influenced by the subjective difference in loudness. If the temporal dimension had been determined by the filled interval, the temporal summation effect would have made the 200-ms interval subjectively louder than the 50-ms interval, and the marmosets might have relied on intensity to discriminate between the 50-ms and 200-ms intervals.

3.4. Procedure

A daily session consisted of 60 trials (4 stimuli × 15 trials). The order of the stimulus presentation was pseudo-randomly determined, such that each of the four stimuli was presented once over the span of four successive trials. The trials began when the marmoset successfully remained on the stimulus perch for a given time, which varied across trials from 3 to 5 s. Next, one of the four stimuli was presented for up to 5 s or until the marmoset moved to the response perch. If the marmoset left the stimulus perch before the stimulus was presented, the trial was discontinued, and the same trial was resumed when the marmoset again stepped onto the stimulus perch.

A GO response to the response perch during the presentation of S^+ resulted in a 1-s access period to a food reward (a HIT), whereas a NOGO response was classified as a MISS. During the presentation of negative stimulus (S–), a GO response was treated as a false alarm (FA) and was followed by a 3-s blackout period. A NOGO response was treated as a correct rejection (CR). CR responses were not followed by explicit rewards. After the MISS and FA trials, the same trial was repeated up to two times to facilitate learning. The results from these correction trials were excluded from the analyses.

For the analysis of the results, first, the GO-response rate in each stimulus condition was determined by random resampling with replacement from trial data (1 for a GO response and 0 for a NOGO response). Then, the d' score was calculated from the HIT and FA rates to determine the discriminability between S^+ and each S–. This procedure was repeated 1000 times to compute the mean of resampled HIT and FA rates and the mean d' scores. A learning criterion for the discriminability of S^+ from each S– was set at a d' score > 1.0. Second, trial data were also resampled by replacing a number of data samples equal to that in each stimulus condition, but the samples were shuffled across conditions to create the expected null distribution of GO-responses. Then, a d' score was calculated according to the contrast between the HIT rate and GO-response rate in the null events to obtain an overall discrimination accuracy to S^+. This procedure was repeated 1000 times to compute the mean and 95% CI of the d' scores. A learning criterion for the overall detectability of S^+ was set at a d' score > 1.0 and a lower bound of 95% CI around $d' > 0$ (Gentner et al., 2006). Training was completed when both criteria were attained for three consecutive sessions. To avoid infinite d' scores, scores of 1 and 0 for HIT and FA rates were transformed to 0.999 and 0.001, respectively.

4. Results and discussion

As shown in Fig. 3A and B, both marmosets learned to suppress responses to S– while maintaining high response rates to S^+. Accordingly, the discrimination between S^+ and each S– gradually improved (Fig. 3C and D). In particular, S^+ (dark solid lines) was efficiently discriminated from S– that differed from S^+ in temporal and frequency dimensions, in Cj190 and Cj195, respectively (gray lines). In Cj195, the classification between responses to S^+ and such S– met the discrimination criterion within ten sessions (Fig. 3D). Eventually, the overall discrimination improved, and Cj190 and Cj195 completed the task in 16 and 24 sessions, respectively. In the last three sessions, the ranges for d' and the lower bound of 95% CI were 1.20–1.73 and 1.15–1.71 in Cj190 and 1.17–1.74 and 1.14–1.71 in Cj195, respectively (Fig. 3E and F).

The results demonstrated that the marmosets were able to discriminate frequencies of 0.5 and 2 kHz and intervals of 50 and 200 ms and also led to two additional findings. First, because compound stimuli were used, the tasks could not be completed if the discrimination relied on one dimension (such as a temporal dimension) or one physical feature (such as a 2 kHz-frequency). Thus, the marmosets were able to combine the frequency and temporal information of sounds. Second, the responses of the marmosets were measured according to absolute discrimination training in which the animals were required to identify the stimuli without any reference; thus, memory and discrimination are involved in selecting the correct response, as previously reported (Elliott and Trahiotis, 1972). Therefore, the marmosets evidently formed a representation of the stimuli in their long-term memory stores.

The temporal extent of the stimuli was determined by an empty interval in Experiment 1 but by a filled duration in Experiments 2 and 3. However, evaluations of the empty interval and filled interval have been found to be equivalent in human subjects in the range used in the current experiments (Hasuo et al., 2011). Therefore, the marmosets may have processed both empty and filled intervals equivalently.

Together, the results of Experiment 1 showed that the marmosets were able to discriminate between the elements that constituted the stimulus patterns used in Experiment 2.

5. Experiment 2: discrimination of sequences: two-frequency × two-duration

Experiment 2 was conducted to determine whether the marmosets were able to perceive structured auditory patterns in terms of the temporal configuration of discrete elements. For this purpose, I trained two marmosets to discriminate between ABAB and AABB patterns; the discriminability of the A and B elements was confirmed in Experiment 1. The discrimination of these patterns requires the ability to process the temporal configuration of individual elements because, similarly to Morse code, one stimulus (e.g., ABAB) is a rearrangement of the elements that also constitute the other stimulus (e.g., AABB). In other words, if the marmosets cannot discriminate between the ABAB and AABB patterns, in which the differences between A and B elements are noticeable, such a failure can be attributed to a difficulty in analyzing the temporal structure of the stimulus patterns.

6. Materials and methods

6.1. Subjects and apparatus

The subjects and apparatus were the same as those used in Experiment 1.

6.2. Stimuli

The stimulus patterns were composed of two common elements in different arrangements. The ABAB and AABB patterns consisted of A (a pure tone lasting 50 ms at 0.5 kHz) and B (a pure tone lasting 200 ms at 2 kHz) elements (Fig. 4). The fade-in and fade-out durations of each

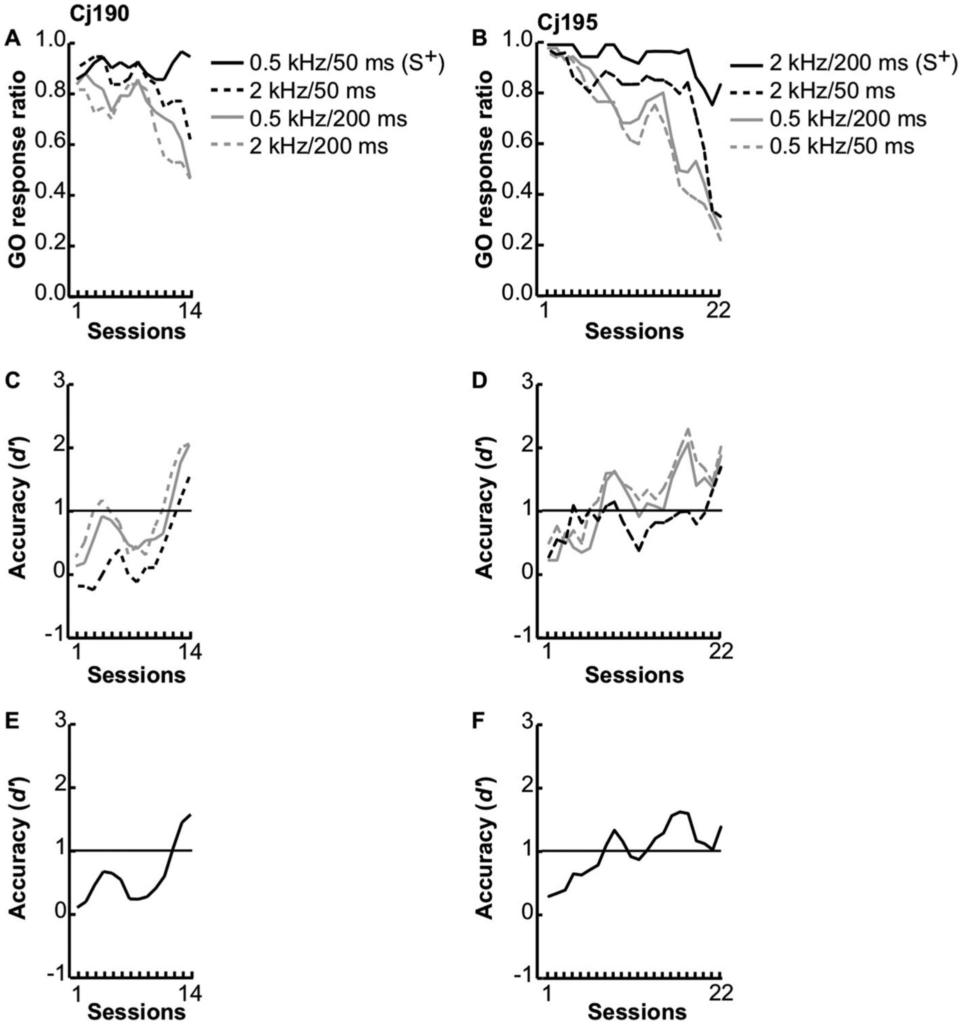


Fig. 3. Results of Experiment 1. (A and B) Plots show mean HIT and FA rates estimated by bootstrapping for Cj190 and Cj195, respectively. (C and D) Plots show mean *d'* scores estimated by bootstrapping the contrast between the HIT rate and FA rate to each S^- for Cj190 and Cj195, respectively. (E and F) Plots show mean d′ scores estimated by bootstrapping the contrast between HIT rate and overall GO-response rate for Cj190 and Cj195, respectively. The plots are averaged over three sessions to eliminate daily fluctuations. The horizontal lines in C–F indicate the learning criterion.

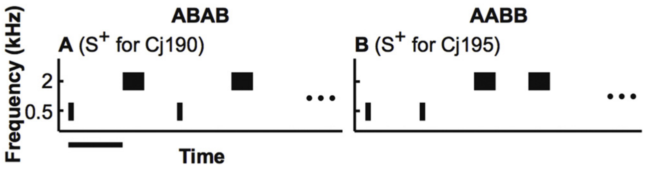


Fig. 4. Schematic representation of the stimuli in Experiment 2. ABAB and AABB patterns were generated by arranging a pure tone of 50-ms duration at a 0.5-kHz frequency and a pure tone of 200-ms duration at a 2-kHz frequency. Stimulus sequences were presented for up to 5 s. Scale bar = 500 ms.

pulse were 10 ms. The physical intensities of the A and B elements were not corrected for [65 dB SPL, 40 dB above the threshold at 0.5 kHz (Osmanski and Wang, 2011)]. Thus, the subjective intensity of B was higher than that of A by approximately 10–15 dB, because marmoset hearing is more sensitive to a 2-kHz tone than a 0.5-kHz tone (Osmanski and Wang, 2011) and because of the probable greater temporal summation effect for a 200-ms tone than for a 50-ms tone. These elements were then combined with a 500-ms onset asynchrony to generate isochronous ABAB and AABB sequences. The B element was more conspicuous than the A element in terms of pitch, duration and possibly intensity; thus, the marmosets might have relied on any one of these dimensions to abstract ABAB and AABB patterns. The stimuli consisted of acoustically simple sounds, and their frequencies were low relative to the range of the marmoset's natural calls (such as distress and mobbing calls). Therefore, the stimuli were unlikely to influence marmoset behavior in terms of ethological relevancy.

Newport et al. (2004) have presented stimulus elements at a more rapid rate; however, the onset asynchrony was not substantially different from those in many previous studies (Fitch and Hauser, 2004; Saffran et al., 2008; Hauser and Glynn, 2009; Ravignani et al., 2013; Wilson et al., 2013, 2015). The duration of the elements was generally longer in previous studies than in the current study. In other words, because the individual elements were separated by a relatively long silence gap in the current experiment, a smooth transition between sounds did not occur; thus, the stimulus sequences may not have provided unexpected prosodic cues. The ABAB and AABB patterns were S^+ for Cj190 and Cj195, respectively.

6.3. Procedure

Experiment 2 started 119 and 9 days after the end of Experiment 1 in Cj190 and Cj195, respectively. The experimental procedure was essentially the same as that in Experiment 1. A daily session consisted of 40 trials (2 stimuli × 20 trials). The order of the S^+ and S– trials was pseudo-randomly determined such that the same stimulus was not presented more than twice in succession. The discrimination achievement and learning criteria were determined in the same manner as in Experiment 1. However, the *d'* scores representing a contrast between S^+ and S– were not calculated because, given that only one S– was presented in Experiment 2, the information that was included in those scores was also indicated by the *d'* scores representing the overall detectability of S^+.

The successful discrimination of auditory patterns achieved after a long period of training by using small sets of exemplars could be because the animals exploited the acoustic features of the stimulus sounds instead of differences in global pattern structures (van Heijningen et al., 2009; ten Cate and Okanoya, 2012).

In Experiment 2, the discrimination achievement did not improve; therefore, training was discontinued at twice the number of sessions required to meet the criteria in Experiment 1.

7. Results and discussion

The two marmosets did not discriminate between the ABAB and AABB patterns within the predetermined training periods. Training was discontinued at 32 and 48 sessions for Cj190 and Cj195, respectively. As shown in Fig. 5A and B, both marmosets maintained a high response rate to S^+ but did not suppress their responses to S–. Consequently, the learning curves indicated no improvement in overall discrimination performance within the studied period (Fig. 5C and D). In the last three sessions, the ranges for *d'* and the lower bound of 95% CI were -0.09 to 0.83 and -0.12 to 0.80 in Cj190 and –0.12 to 0.20 and –0.15 to 0.16 in Cj195, respectively.

The results indicated that the marmosets have difficulty in processing the temporal structure of stimulus patterns. In Experiment 2, because the B element was higher in frequency, longer in duration and subjectively louder than the A element, the B element was more conspicuous than the A element. Therefore, the marmosets may have efficiently abstracted ABAB and AABB patterns regardless of the dimensions relied on to analyze the temporal structure of the stimulus patterns. However, the marmosets were unable to discriminate between ABAB and AABB patterns.

One potential explanation for why the marmosets did not attain discrimination is that the marmosets were not deprived of food during the training period, and their motivation may have been insufficient for the demanding task. However, high response rates (at least to S^+) were always observed until the end of each session, thus indicating that the reinforcement value of the reward was sufficiently high to maintain the monkeys' behavior during the daily training and that food restriction was not critical in the current studies.

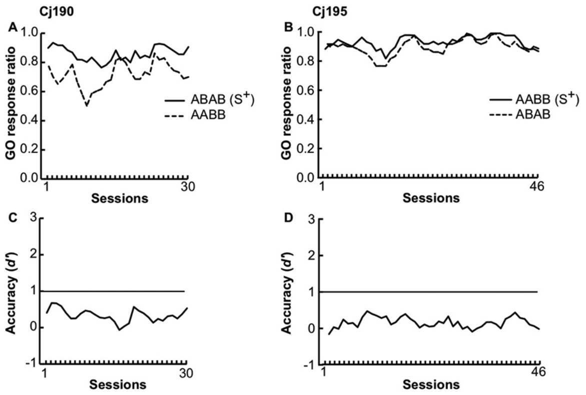


Fig. 5. Results of Experiment 2. (A and B) Plots show mean HIT and FA rates estimated by bootstrapping for Cj190 and Cj195, respectively. (C and D) Plots show mean *d'* scores estimated by bootstrapping the contrast between HIT rate and overall GO-response rate for Cj190 and Cj195, respectively. For plot detail, see legend to Fig. 3.

In addition, the learning history during Experiment 1 might have caused confusion regarding the stimulus patterns in Experiment 2. Both ABAB and AABB patterns contained acoustic features that were S^+ in Experiment 1 (0.5-kHz/50-ms feature in Cj190 and 2-kHz/200-ms feature in Cj195, corresponding to A and B elements, respectively). However, temporal information was provided by the silent interval between elements, and the duration of the elements was not task relevant in Experiment 1, whereas the duration of each element provided task-relevant temporal information in Experiment 2. Moreover, the start of Experiment 2 was separated from the end of Experiment 1 by a long period after the end of Experiment 1. Therefore, it is unlikely that confusion in Experiment 2 was induced by the learning history during Experiment 1.

Instead, the results of Experiment 2 may have been caused by the structural nature of the stimuli. Because the stimulus patterns were fixed, and stimuli were initiated by the marmosets' own responses to the stimulus perch, the marmosets might have noticed the difference between the ABAB and AABB patterns if they had been able to detect whether the second tone was A or B. Moreover, by ignoring the less conspicuous A element, the marmosets could have compared the "-B-B" pattern, in which the 2-kHz/200-ms tone was regularly repeated with a 1000-ms onset asynchrony, with the "—BB" pattern, in which the 2-kHz/200-ms tone was repeated with either a 500- or 1500-ms onset asynchrony. If this were the case, these interval differences may indeed have been noticeable for the marmosets and allowed them to successfully discriminate between the stimulus patterns. Thus, the failure to discriminate between the ABAB and AABB patterns may indicate that the marmosets' judgment of the stimuli did not rely on single elements. In other words, the failure in temporal pattern discrimination was likely to be dependent on the structural nature of the stimuli.

8. Experiment 3: discrimination of sequences: two-frequency × one-duration (or one-frequency × two-duration)

The results in Experiment 2 were in contrast to those in Experiment 1, in which identical tones were regularly repeated in each stimulus. The marmosets were unable to discriminate the stimuli, probably because they had difficulty in processing the temporal configuration of discrete elements to represent the temporal structure of stimulus patterns.

Therefore, in Experiment 3, to confirm that this failure was due to the structural nature of the stimuli, I trained the marmosets to discriminate four stimulus patterns. In each stimulus pattern, either the frequency or the temporal dimension was composed of two elements, whereas the other dimension was composed of a single element. Thus, the structured patterns were generated in only one dimension. Therefore, if marmosets confused the duration-based (or frequency-based) ABAB and AABB patterns, but their responses were determined by the frequency-based (or duration-based) temporally non-structured features, then the marmosets would show difficulty in representing the temporal structure of the stimulus patterns.

9. Materials and methods

9.1. Subjects and apparatus

The subjects and apparatus were the same as those in Experiments 1 and 2.

9.2. Stimuli

Two marmosets were trained with different stimulus sets. Cj190 was exposed to four stimuli in which 50- (A) and 200-ms (B) tones were arranged to generate ABAB and AABB patterns. The frequency of the elements within each stimulus was constant and was either 0.5 or 2 kHz (Fig. 6A–D). Cj190 was trained with a duration-based ABAB pattern at

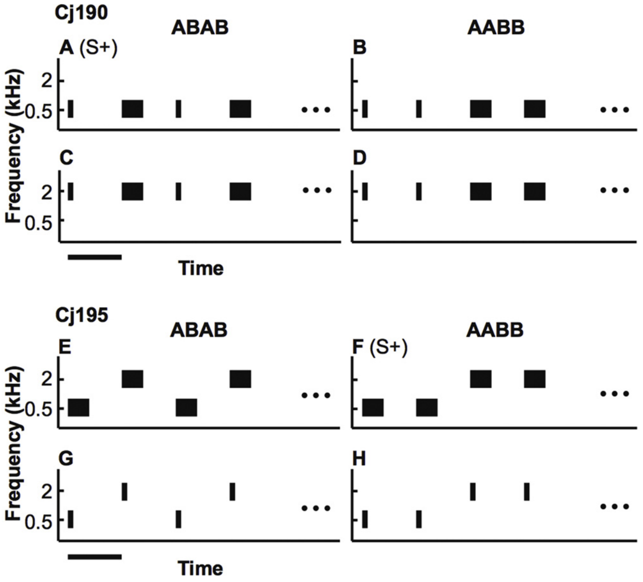


Fig. 6. Schematic representation of the stimuli in Experiment 3. Cj190 was exposed to four stimuli in which 50- and 200-ms tones were arranged to generate ABAB (A and C) and AABB (B and D) patterns during phase 1. The frequencies of the elements within the individual stimuli were identical and were either 0.5 (A and B) or 2 kHz (C and D). During phase 2, two stimulus patterns (A and B) were presented. Cj195 was exposed to four stimuli in which 0.5- and 2-kHz tones were arranged to generate ABAB (E and G) and AABB (F and H) patterns during phase 1. The duration of the elements within the individual stimuli was identical and was either 50 (G and H) or 200 ms (E and F). During phase 2, two stimulus patterns (E and F) were presented. The stimulus sequences were presented for up to 5 s. Scale bar = 500 ms.

the 0.5-kHz frequency as S^+ (Fig. 6A). Thus, the temporal and frequency dimensions constituted structured and non-structured features, respectively.

In contrast, Cj195 was exposed to four stimuli in which 0.5- (A) and 2-kHz (B) tones were arranged to generate ABAB and AABB patterns. The duration of the elements within each stimulus was constant and was either 50 or 200 ms (Fig. 6E–H). Cj195 was trained by using the frequency-based AABB pattern with a 200-ms duration as S^+ (Fig. 6F). Thus, the frequency and temporal dimensions constituted structured and non-structured features, respectively.

9.3. Procedure

The procedure was essentially the same as that in Experiment 2. However, Experiment 3 consisted of two phases. Phase 1 started four days after the end of Experiment 2 in both Cj190 and Cj195. A daily session consisted of 60 trials (4 stimuli × 15 trials). The order of stimulus presentation was pseudo-randomly determined, such that each of the four stimuli was presented once in four trials. The discrimination performance was calculated, and the learning criterion was set in the same manner as in Experiment 1. However, because there was no improvement in achievement (as in Experiment 2), training was discontinued at twice the number of sessions required to achieve the task in Experiment 1 (32 and 48 sessions for Cj190 and Cj195, respectively).

Additional training (phase 2) was conducted by eliminating possible interference from a non-structured feature and started 60 and 4 days after the end of phase 1 in Cj190 and Cj195, respectively. Thus, with two stimulus sequences, the marmosets were trained to discriminate between ABAB and AABB patterns that shared the non-structured features of S^+ (Fig. 6A and B for Cj190 and Fig. 6E and F for Cj195). A daily session consisted of 40 trials (2 stimuli × 20 trials). The order of stimulus presentation was pseudo-randomly determined, such that the same stimulus was not presented more than twice in succession. Consequently, there was no improvement in achievement; training was thus

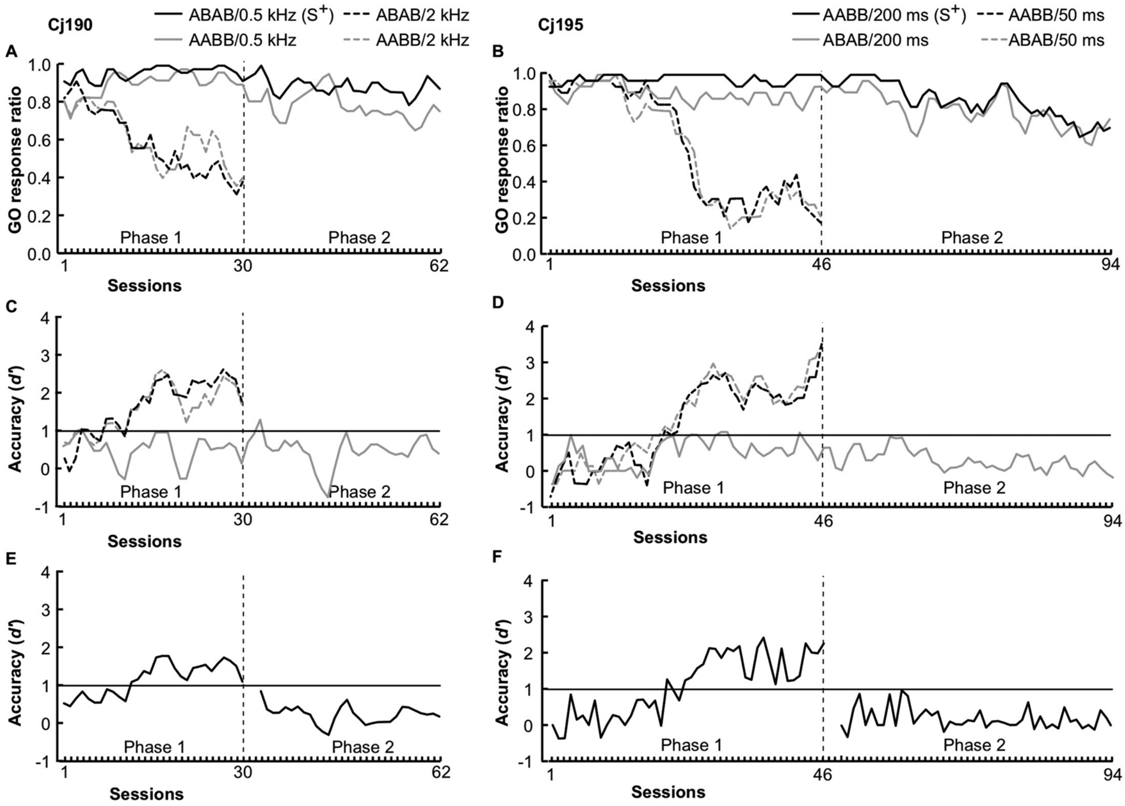


Fig. 7. Results of Experiment 3. See legend to Fig. 3.

discontinued at twice the number of sessions required to achieve the task in Experiment 1 (32 and 48 sessions for Cj190 and Cj195, respectively).

10. Results and discussion

In phase 1, as shown in Fig. 7A and B, both marmosets learned to inhibit their responses to two S– that did not share the non-structured feature with S^+ (dashed lines), while maintaining their response to S^+ (dark solid lines). However, they did not suppress their responses to S– that shared the non-structured feature with S^+ (gray solid lines). In other words, their discrimination relied on the non-structured feature across sessions.

As shown in Fig. 7C and D, discrimination improved between S^+ and S– that did not share the non-structured feature (dashed lines) of frequency in Cj190 and duration in Cj195. Cj190 and Cj195 attained the discrimination criterion between S^+ and these S– within 18 and 25 sessions, respectively. However, S^+ was not discriminated from S– that differed in the structured feature (gray solid lines) within the training period. In the last three sessions of phase 1, ranges for *d′* and the lower bound of 95% CI were –0.46 to 0.45 and –0.51 to 0.41 in Cj190 and 0.03 to 0.84 and –0.02 to 0.80 in Cj195, respectively (Fig. 7C and D). Consequently, although the overall response accuracy appeared to increase across sessions (Fig. 7E and F), this change merely reflected a partial discrimination.

In phase 2, both marmosets consistently had similar numbers of GO responses to both S^+ and S– (Fig. 7A and B). Consequently, discrimination did not improve (Fig. 7C–F). Thus, the results of Experiment 3 again showed that the marmosets were unable to discriminate the stimulus patterns within the training period. In the last three session of phase 2, the ranges for *d′* and the lower bound of 95% CI were 0.01–0.81 and –0.02 to 0.84 in Cj190 and –0.36 to 0.13 and –0.39 to 0.11 in Cj195, respectively (Fig. 7E and F).

The marmosets' behavior in phase 1, together with the results in Experiment 1, indicated that their stimulus judgment relied on non-structured features. In phase 2, in which the structured feature was the only cue for discrimination, the marmosets did not correctly choose S^+. Thus, together with the results of Experiment 2, the results of phase 2 indicated that the marmosets were unable to process the temporal configuration of discrete auditory elements. In summary, marmosets have difficulty in perceiving the temporal structure of auditory patterns.

11. General discussion

The present study was aimed at clarifying whether marmosets can perceive auditory patterns in terms of the temporal configuration of elements rather than the acoustic sensory features. To this end, marmosets were trained to discriminate auditory temporal patterns in which two discriminable elements were arranged in a different order. Thus, the marmosets had to analyze the temporal configuration of elements to perceive tonal patterns. The results showed that the marmosets were able to recognize the acoustic features of stimuli but not to discriminate the patterned structures of stimuli. These findings were obtained from only two marmosets; however, the results from one animal were replicated by those from the other animal. Thus, the present study showed that common marmosets cannot perceive unified auditory patterns on the basis of the temporal configuration of discrete tones.

Several possible explanations should be ruled out before reaching a conclusion. First, one possibility is that the marmosets may have failed to achieve the tasks because they were not trained sufficiently. In Experiments 2 and 3, Cj190 and Cj195 were exposed to S^+ for 1760 and 2640 trials, respectively. They might have achieved the tasks in Experiments 2 and 3 if they had been trained for a longer period. European starlings have successfully learned to discriminate differently structured auditory sequences after approximately 10,000–50,000 trials, in which each stimulus was presented a few thousand times (Gentner et al., 2006). However, van Heijningen et al. (2009) and ten Cate and Okanoya (2012) have noted that the birds might have exploited unexpected acoustic features of the stimulus patterns and discriminated the stimuli on the basis of counting the repetition of elements belonging to the stimulus sequences of the same category. Thus, it was likely that a perceptual learning process to select the sensory features distinguishing S^+ and S–, rather than learning of the structural rules of the stimulus sequences, required a long training period. As Brosch et al. (2006) have noted, macaque monkeys have learned to discriminate tone sequences after more than 100,000 trials (Brosch et al., 2004), whereas a few thousand trials were required to achieve a similar task by Izumi (2001). Thus, training length may not necessarily determine learning performance. If marmosets can process structured auditory patterns on the basis of the configuration of individual elements, lengthy training may therefore not have been necessary to complete the current temporal pattern discrimination. Nevertheless, it may be necessary to test whether a long period of training results in better discrimination performance.

A second possibility is that the marmosets may have been unable to discriminate auditory patterns in Experiments 2 and 3 because the training history in Experiment 1, in which identical tones were monotonously repeated, may have prevented the marmosets from attending to the temporal configuration of discrete tones. In the case of Cj190, for example, an element that organized S^+ of Experiment 1 (i.e., 0.5-kHz and 50-ms) was included in both stimuli in Experiment 2 and in the two stimuli in Experiment 3, to which that marmoset showed a sustained response. Thus, it can be assumed that the marmosets did not discriminate the stimuli because they were persistently responding to the element in the later training sessions. However, this possibility is unlikely because the start of Experiment 2 was separated from the end of Experiment 1 by a long period (roughly four months) after the end of Experiment 1 for this marmoset, and comparable results were obtained from the other marmoset. In Experiment 1, S^+ for Cj195 was structured by an element that was not contained in the stimuli in Experiment 2 (i.e., 2-kHz and 50-ms). In Experiment 3, moreover, this marmoset successfully avoided stimuli that contained this element. Thus, the results in Experiments 2 and 3 cannot easily be explained in terms of an influence of learning history during Experiment 1.

A third possibility is that the high cognitive load for the task may have prevented the marmosets from achieving discrimination. Critical differences exist between the current study and prior studies suggesting that an ability to notice the differences in the sound sequences in monkeys is also present in the task requirements: the current study was related to training to discriminate auditory sequences, whereas the former studies were related to the detection of stimulus novelty, as measured by spontaneous behavior. In the current experiments, moreover, the stimulus presentations were separated by a silent inter-trial interval; thus, the marmosets were required to judge the stimuli solely on the basis of the absolute features stored in long-term memory. If the target stimuli had been embedded in continuously presented standard stimuli, for instance, the subjects could have detected the target by relying on the relative change in the ongoing auditory input without having to rely on long-term memory. Undoubtedly, the cognitive load for tasks involving absolute discrimination of stimuli is higher than that for tasks requiring only relative discrimination. In future studies, therefore, it will be important to employ discrimination tasks that require a reduced cognitive load, such as the detection of stimulus changes for common marmosets. Comparisons of results from tasks requiring different degrees of cognitive load should be helpful in clarifying the features of auditory sequence processing in non-human primates.

Fourth, the stimulus configuration might have caused a potential problem. Because the stimulus patterns were fixed during the training sessions, stimuli might have provided stimulus-unique local transitions as cues for discrimination without requiring the marmosets to attend to the entire structure of the stimuli. The marmosets may have noticed the difference between the stimulus patterns if they had been able to detect the stimulus-unique transitions (i.e., AA and BB). However, because both sequences were isochronous in the present study, the stimulus-unique features might have been obscure. If such features had been less conspicuous than AB and BA transitions, which were included in both stimulus sequences, the marmosets would not have noticed the difference between S^+ and S–. Therefore, future studies may use stimuli composed of the same elements but arranged with biased onset asynchrony, such that the marmosets can efficiently make a perceptual grouping (i.e., AB-AB vs. AA-BB patterns), to study whether marmosets can discriminate stimulus patterns if element configurations are modified.

Although there are several limitations, as mentioned above, the current experiments at least preliminarily revealed monkeys' limited ability to learn auditory patterns. To explain the current results, cognitive mechanisms essential for processing the temporal configuration of sounds may be worthwhile to consider. To complete the current discrimination tasks (Experiments 2 and 3), the marmosets were required to keep track of sequentially presented auditory elements and to maintain the temporal configuration of elements in short-term memory before they perceived the structures of the ABAB and AABB patterns. Monkeys can hold information on a single tone in short-term memory for longer than 10 s (D'Amato and Colombo, 1985; Kojima, 1985; Colombo and D'Amato, 1986; Fritz et al., 2005). However, Izumi (1999) has shown that a brief silence gap inserted between tones influences the judgement of the pitch relation. Additionally, Scott and colleagues have suggested that the memory of a preceding sound can be overwritten by an incoming sound (Scott et al., 2012). Furthermore, Wright (2002) has demonstrated that the memory of a preceding sound can also influence the memory of incoming sounds. In the current experiments, therefore, the marmosets were unable to discriminate between the ABAB and AABB patterns, probably because the processing of the preceding and incoming A/B elements mutually interfered, thus resulting in an unstructured perception of the element sounds.

Although I did not directly survey neural substrates for auditory sequence processing in this work, recent comparative studies of brain function coupled with habituation/dishabituation methods may be helpful to explain the present results. An electrophysiological study (Milne et al., 2016) has revealed that contrasts in event-related potential patterns evoked by congruent and incongruent sequences from monkeys are similar to those from human infants rather than human adults. Thus, such findings indicate similar strategies for detecting rule deviations in sound sequences in both monkeys and human infants. Therefore, the temporal configuration of individual tones may not be an important cue for perception of auditory sequences in monkeys, because infants' recognition of rule-based auditory sequences is dependent on pitch changes or prosodic information (Mueller et al., 2012; Shukla et al., 2011; for reviews, see Jusczyk, 1999; Erickson and Thiessen, 2015). Furthermore, by functional brain imaging, Wang et al. (2015) have found brain regions (i.e., bilateral inferior frontal and superior temporal regions) that represent the rule of global sequence patterns in humans, but no such regions have been found in monkeys, although only some aspects of sequence were represented.

Recent comparative studies of brain anatomy may also be helpful in explaining the present results. In humans, the left posterior part of the inferior frontal region, i.e. Brodmann areas (BA) 44 and 45, is essential for sequencing phonemes and words, thereby enabling the detection of words from continuous speech and the analysis of grammatical structures. This region receives auditory signals from the temporal auditory regions via the arcuate fasciculus of the dorsal pathway (for reviews, see Dick and Tremblay, 2012; Friederici et al., 2017). However, left BAs 44 and 45 are more than six times smaller in chimpanzee than humans, while the brain as a whole is roughly 3.5 times smaller in chimpanzees than humans (Schenker et al., 2010). In addition, the arcuate fasciculus of macaques and chimpanzees is rudimentary compared with that in humans (for reviews, see Petrides et al., 2012; Rilling, 2014). These comparative studies of brain functions and structures imply that macaques and chimpanzees, which are phylogenetically closer to humans than marmosets, do not have sufficient neural substrates that support auditory sequence processing based on the temporal configuration of sound. In the current experiments, therefore, the marmosets were unable to discriminate between the ABAB and AABB patterns, probably because they do not have neural substrates for perceiving unified auditory patterns. Although caution must be taken in generalizing the current negative results, the findings from previous studies support the speculation that auditory temporal pattern perception may be difficult not only in marmosets but also in other non-human primates.

In conclusion, the present study showed that common marmosets can recognize the physical properties of sound but cannot represent the structure of auditory patterns by coding the temporal configuration of individual sounds. The current results can be explained by the nature of auditory cognition and brain mechanisms in monkeys. Monkeys may represent tone sequences as unstructured pools of individual tones, and their brains may not represent the global structure of auditory temporal patterns. Owing to such limitations, therefore, even if the sequential stimuli had been composed of acoustically complex sounds or species-specific calls, which should attract more attention than pure tones, the marmosets would have failed to discriminate between the auditory patterns. The current findings were not consistent with previous findings suggesting a sensitivity to the structural rules of sound sequences in monkeys. These discrepant findings were probably due to the differences in the experimental procedures. Previous studies have typically measured spontaneous behavioral responses and evoked brain activations against passively exposed sound stimuli. Although different task procedures have their benefits and drawbacks, in future studies, carefully controlled behavioral training with sophisticated procedures will be needed to confirm the auditory sequence perception abilities of nonhuman primates.

Conflicts of interest

The author declares no conflict of interest related to this manuscript.

Acknowledgments

The study was supported by a Grant-in-Aid for Challenging Exploratory Research [JSPS KAKENHI 26540066]. I thank Mitsue Wakita for preparing the marmosets' food pellets, as well as the two anonymous reviewers for their contributions in improving this paper.

References

Beckers, G.J.L., Berwick, R.C., Okanoya, K., Bolhuis, J.J., 2017. What do animals learn in artificial grammar studies? Neurosci. Biobehav. Rev. 81, 238–246.

Bezerra, B.M., Souto, A., 2008. Structure and usage of the vocal repertoire of *Callithrix jacchus*. Int. J. Primatol. 29, 671–701.

Brosch, M., Selezneva, E., Bucks, C., Scheich, H., 2004. Macaque monkeys discriminate pitch relationships. Cognition 91, 259–272.

Brosch, M., Oshurkova, E., Bucks, C., Scheich, H., 2006. Influence of tone duration and intertone interval on the discrimination of frequency contours in a macaque monkey. Neurosci. Lett. 406, 97–101.

Colombo, M., D'Amato, M.R., 1986. A comparison of visual and auditory short-term memory in monkeys (*Cebus paella*). Q. J. Exp. Psychol. B 38, 425–448.

D'Amato, M.R., Colombo, M., 1985. Auditory matching-to-sample in monkeys (*Cebus paella*). Anim. Learn. Behav. 13, 375–382.

D'Amato, M.R., Colombo, M., 1988. On tonal pattern perception in monkeys (*Cebus apella*). Anim. Learn. Behav. 16, 417–424.

D'Amato, M.R., Salmon, D.P., 1984. Processing of complex auditory stimuli (tunes) by rats and monkeys (*Cebus apella*). Anim. Learn. Behav. 12, 184–194.

Dick, A.S., Tremblay, P., 2012. Beyond the arcuate fasciculus: consensus and controversy in the connectional anatomy of language. Brain 135, 3529–3550.

Elliott, D.N., Trahiotis, C., 1972. Cortical lesions and auditory discrimination. Psychol. Bull. 77, 198–222.

Erickson, L.C., Thiessen, E.D., 2015. Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. Dev. Rev. 37, 66–108.

Fitch, W.T., Hauser, M.D., 2004. Computational constraints on syntactic processing in a nonhuman primate. Science 303, 377–380.

Friederici, A.D., Chomsky, N., Berwick, R.C., Moro, A., Bolhuis, J.J., 2017. Language, mind and brain. Nat. Hum. Behav. 1, 713–722.

Fritz, J., Mishkin, M., Saunders, R., 2005. In search of an auditory engram. Proc. Natl. Acad. Sci. U. S. A. 102, 9359–9364.

Gentner, T.Q., Fenn, K.M., Margoliash, D., Nusbaum, H.C., 2006. Recursive syntactic pattern learning by songbirds. Nature 440, 1204–1207.

Ghirlanda, S., 2017. Can squirrel monkeys learn an AB^nA grammar? A re-evaluation of Ravignani et al. (2013). PeerJ 5, e3806.

Hasuo, E., Nakajima, Y., Ueda, K., 2011. Does filled duration illusion occur for very short time intervals? Acoust. Sci. Technol. 32, 82–85.

Hauser, M.D., Glynn, D., 2009. Can free-ranging rhesus monkeys *(Macaca mulatta)* extract artificially created rules comprised of natural vocalizations? J. Comp. Psychol. 123, 161–167.

Izumi, A., 1999. The effect of marker frequency disparity on the discrimination of gap duration in monkeys. Perception 28, 437–444.

Izumi, A., 2001. Relative pitch perception in Japanese monkeys (*Macaca fuscata*). J. Comp. Psychol. 115, 127–131.

Jusczyk, P.W., 1999. How infants begin to extract words from speech. Trends Cogn. Sci. 3, 323–328.

Kojima, S., 1985. Auditory short-term memory in the Japanese monkey. Int. J. Neurosci. 25, 255–262.

Milne, A.E., Mueller, J.L., Männel, C., Attaheri, A., Friederici, A.D., Petkov, C.I., 2016. Evolutionary origins of non-adjacent sequence processing in primate brain potentials. Sci. Rep. 6, 36259.

Mueller, J.L., Friedericim, A.D., Männel, C., 2012. Auditory perception at the root of language learning. Proc. Natl. Acad. Sci. U. S. A. 109, 15953–15958.

Newport, E.L., Hauser, M.D., Spaepen, G., Aslin, R.N., 2004. Learning at a distance II. Statistical learning of non-adjacent dependencies in a non- human primate. Cogn. Psychol. 49, 85–117.

Osmanski, M.S., Wang, X., 2011. Measurement of absolute auditory thresholds in the common marmoset (*Callithrix jacchus*). Hear. Res. 277, 127–133.

Petrides, M., Tomaiuolo, F., Yeterian, E.H., Pandya, D.N., 2012. The prefrontal cortex: comparative architectonic organization in the human and the macaque monkey brains. Cortex 48, 46–57.

Pistorio, A.L., Vintch, B., Wang, X., 2006. Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). J. Acoust. Soc. Am. 120, 1655–1670.

Ravignani, A., Sonnweber, R.S., Stobbe, N., Fitch, W.T., 2013. Action at a distance: dependency sensitivity in a New World primate. Biol. Lett. 9, 20130852.

Rilling, J.K., 2014. Comparative primate neurobiology and the evolution of brain language systems. Curr. Opin. Neurobiol. 28, 10–14.

Saffran, J.R., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., Cushman, F., 2008. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. Cognition 107, 479–500.

Schenker, N.M., Hopkins, W.D., Spocter, M.A., Garrison, A.R., Stimpson, C.D., Erwin, J.M., Hof, P.R., Sherwood, C.C., 2010. Broca's area homologue in chimpanzees (*Pan troglodytes*): probabilistic mapping, asymmetry, and comparison to humans. Cereb. Cortex 20, 730–742.

Scott, B.H., Mishkin, M., Yin, P., 2012. Monkeys have a limited form of short-term memory in audition. Proc. Natl. Sci. U. S. A. 109, 12237–12241.

Shukla, M., White, K.S., Aslin, R.N., 2011. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. Proc. Natl. Acad. Sci. U. S. A. 108, 6038–6043.

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

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

Wang, L., Uhrig, L., Jarraya, B., Dehaene, S., 2015. Representation of numerical and sequential patterns in macaque and human brains. Curr. Biol. 25, 1966–1974.

Wilson, B., Slater, H., Kikuchi, Y., Milne, A.E., Marslen-Wilson, W.D., Smith, K., Petkov, C.I., 2013. Auditory artificial grammar learning in macaque and marmoset monkeys. J. Neurosci. 33, 18825–18835.

Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., Thiele, A., Griffiths, T.D., Marslen-Wilson, W.D., Petkov, C.I., 2015. Auditory sequence processing engages evolutionarily conserved regions of frontal cortex in macaques and humans. Nat. Commun. 6, 8901.

Wright, A.A., 2002. Monkey auditory list memory: test with mixed and blocked retention delays. Anim. Learn. Behav. 30, 158–164.

Wright, A.A., Rivera, J.J., Hulse, S.H., Shyan, M., Neiworth, J.J., 2000. Music perception and octave generalization in rhesus monkeys. J. Exp. Psychol. Gen. 129, 291–307.