LIMITING PARENTAL INTERACTION DURING VOCAL DEVELOPMENT AFFECTS ACOUSTIC CALL STRUCTURE IN MARMOSET MONKIES

Yasemin B. Gultekin and Steffen R. Hage*

INTRODUCTION

Human vocal development is dependent on learning by imitation through social feedback between infants and caregivers. Recent studies have revealed that vocal development is also influenced by parental feedback in marmoset monkeys, suggesting vocal learning mechanisms in nonhuman primates. Marmoset infants that experience more contingent vocal feedback than their littermates develop vocalizations more rapidly, and infant marmosets with limited parental interaction exhibit immature vocal behavior beyond infancy. However, it is yet unclear whether direct parental interaction is an obligate requirement for proper vocal development because all monkeys in the aforementioned studies were able to produce the adult call repertoire after infancy. Using quantitative measures to compare distinct call parameters and vocal sequence structure, we show that social interaction has a direct impact not only on the maturation of the vocal behavior but also on acoustic call structures during vocal development. Monkeys with limited parental interaction during development show systematic differences in call entropy, a measure for maturity, compared with their normally raised siblings. In addition, different call types were occasionally uttered in motif-like sequences similar to those exhibited by vocal learners, such as birds and humans, in early vocal development. These results indicate that a lack of parental interaction leads to long-term disturbances in the acoustic structure of marmoset vocalizations, suggesting an imperative role for social interaction in proper primate vocal development.
behavior in marmosets, suggesting that social reinforcement is actively guiding marmoset vocal development (12). Here, we investigate whether parental interaction is crucial for the proper development of the acoustic call structure by comparing the acoustic structures of distinct marmoset call types during the subadult stage. We used two sets of offspring from the same parents that were used in our previous study (12): One set was normally raised, whereas the other was separated from their parents after the third postnatal month. Using quantitative measures to compare distinct call parameters, such as call duration, frequency, and entropy, we show that social feedback has a direct impact on the acoustic call structure during vocal development. Monkeys with limited parental interaction during vocal development show systematic differences in call entropy, that is, vocal immaturity, compared with their normally raised siblings, which cannot be simply explained by physical growth. In addition, our data support a recently proposed model, suggesting a transformation from high-entropy, immature calls to low-entropy, mature vocalizations. Moreover, we observed that during babbling-like behavior, calls were not always uttered in a random order but were occasionally emitted in motif-like sequences similar to those exhibited by vocal learners such as songbirds during early vocal development (4). Moreover, within those sequences, the cries were occasionally exchanged with trills or subharmonic phees within these motif-like sequences, indicating a similar usage of these call types within babbling-like bouts.

RESULTS

Here, we analyzed 8531 calls from five subadult marmoset monkeys in a directed context, that is, with visual and auditory contact with potential caregivers. The subjects were born from the same parents in two different litters (Fig. 1A). The first litter (S1) consisted of one male (S1-3), which was hand-raised by an animal caretaker. Nevertheless, it had several hours of daily auditory and visual contact with its parents and two siblings (S1-1 and S1-2), which were separated from their parents at the age of 3 months to produce a stable social group with the hand-raised sibling. Therefore, all S1 animals lacked direct parental interaction after the third postnatal month (limited parental interaction siblings). The second litter (S2) consisted of a female (S2-1) and a male (S2-2) that were still grouped with the parents at the time of the vocal recordings (normally raised siblings). At the time of recording, both S1 and S2 monkeys were in the subadult stage (Fig. 1A). The S1 monkeys were 13 months old and weighed 380, 400, and 430 g, whereas the S2 siblings were 7 months old and weighed 290 g each. Body weight is a suitable proxy for overall growth as it strongly correlates with vocal apparatus size in monkeys. We hypothesized that if changes in the acoustic call parameters are completely due to physical maturation without any influence of parental interaction, then the slightly older and therefore heavier siblings (S1) should—if changes do occur—produce more mature calls that are lower in call frequency (17) and entropy (11, 14). With respect to call duration, trill and twitter calls should not systematically differ between S1 and S2 siblings, because their duration does not change markedly during vocal development and rather reflects interindividual differences (17). In contrast, the phee calls of the older and heavier S1 siblings should be longer in duration compared with their younger and lighter S2 siblings (17).

Comparison of distinct call parameters in different call types of S1 and S2 monkeys

As a first step, we evaluated the distribution of distinct acoustic call parameters, such as call duration, frequency, and Wiener entropy, of the major adult call types that were frequently produced by all five animals: phee, trill, and twitter calls (Fig. 1B). We analyzed the distributions of these call parameters for each call type separately rather than pooling all uttered vocalizations to identify potential differences and variability within acoustic features between these call types. We observed that all three call parameters were significantly different between monkeys for all tested call types, that is, trill ($P_{\text{dur}} = 5.9 \times 10^{-188}$, $P_{\text{freq}} = 1.5 \times 10^{-156}$, $P_{\text{entropy}} = 7.6 \times 10^{-167}$), $\chi^2_{\text{dur}} = 874.4$, $\chi^2_{\text{freq}} = 777.0$, $\chi^2_{\text{entropy}} = 729.4$, five animals, $n = 5680$ vocalizations, Kruskal-Wallis test), twitter
(P_dur = 1.4 \times 10^{-77}, P_{freq} = 2.6 \times 10^{-55}, P_{entropy} = 1.6 \times 10^{-24}, \chi^2_{dur} = 364.3, \chi^2_{freq} = 117.8, \chi^2_{entropy} = 261.1, five animals, n = 1816 vocalizations, Kruskal-Wallis test), and phee calls (P_dur = 1.2 \times 10^{-139}, P_{freq} = 1.2 \times 10^{-40}, P_{entropy} = 5.9 \times 10^{-121}, \chi^2_{dur} = 651.3, \chi^2_{freq} = 565.0, \chi^2_{entropy} = 2210, five animals, n = 1001 vocalizations, Kruskal-Wallis test). Inter-individual differences in call structure are quite common between individuals of one species (9, 12, 17). Therefore, we performed a post hoc pairwise analysis of the parameter distributions of individuals for all three call types between the S1 and S2 litter to test for systematic differences in call duration, frequency, and Wiener entropy (post hoc multiple comparison tests).

**Trill calls**

We found no systematic distribution of call duration or entropy in either the S1 or the S2 siblings (Fig. 1C). Although one S1 monkey (S1-1) showed a significantly longer median call duration than monkey S2-2 (P = 5.6 \times 10^{-11}, post hoc multiple comparison test), the other two S1 monkeys (S1-2 and S1-3) produced significantly shorter trills than their S2 siblings (P = 4.8 \times 10^{-11} and P = 1.2 \times 10^{-11}, respectively, post hoc multiple comparison test). We also found that the call entropies for both litters were clustered together. One S1 monkey (S1-3) called with significantly higher entropy than that of the S2 group (P = 4.5 \times 10^{-3} and P = 4.9 \times 10^{-17}, post hoc multiple comparison test), whereas its two littermates (S1-1 and S1-2) exhibited significantly lower Wiener entropy measures than the normally raised S2 monkeys (S1-1: P = 1.3 \times 10^{-13} and P = 1.3 \times 10^{-25}; S1-2: P = 6.6 \times 10^{-3} and P = 2.8 \times 10^{-3}; post hoc multiple comparison test). However, with respect to call frequency, we observed that two S1 monkeys uttered calls with significantly lower call frequencies compared with the S2 animals (S1-1: P = 8.1 \times 10^{-22} and P = 1.2 \times 10^{-22}; S1-2: P = 3.7 \times 10^{-9} and P = 0.03; post hoc multiple comparison test), with one S1 monkey (S1-3) showing a lower frequency distribution than S2-1 (P = 4.2 \times 10^{-7}, post hoc multiple comparison test) and a similar distribution to S2-2 (P = 1, post hoc multiple comparison test).

**Twitter calls**

Similar to the trill call findings, nonsystematic distributions were also found for call duration and call frequency within the twitter calls of the S1 and S2 monkeys (Fig. 1D). In general, twitter syllables are very short and the S1 monkeys’ twitter calls could not be clearly distinguished from those of their S2 siblings. Whereas one S2 monkey (S2-1) produced significantly longertwitter syllables than all S1 monkeys (P = 6.1 \times 10^{-30}, P = 4.3 \times 10^{-32}, and P = 7.7 \times 10^{-38}, post hoc multiple comparison test), two S1 monkeys (S1-2 and S1-3) additionally exhibited significantly shorter syllables than those of S2-2 (P = 4.2 \times 10^{-5} and P = 3.5 \times 10^{-19}, post hoc multiple comparison test). One S1 monkey and one S2 monkey marginally did not show any differences in call duration distribution (P = 0.052, post hoc multiple comparison test). For call frequency, one S1 monkey (S1-3) produced significantly lower frequencies than both S1 monkeys (P = 9.3 \times 10^{-22} and P = 3.3 \times 10^{-3}, post hoc multiple comparison test); the two other S1 siblings (S1-1 and S1-2) produced significantly lower frequencies than monkey S2-1 (P = 4.3 \times 10^{-14} and P = 2.1 \times 10^{-3}, post hoc multiple comparison test). Regarding call entropy, all S1 and S2 animals clearly differed from each other. Here, all S1 animals produced twitter calls with significantly lower call entropy distributions than their S2 siblings (all S1 versus S2-1: P = 8.4 \times 10^{-34}, P = 1.3 \times 10^{-36}, and P = 3.4 \times 10^{-36}; all S1 versus S2-2: P = 6.2 \times 10^{-11}, P = 2.9 \times 10^{-21}, and P = 2.8 \times 10^{-3}; post hoc multiple comparison test). Twitter calls are regularly produced repetitively often with eight or more consecutive twitter syllables (so-called phrases) (17). Therefore, we also tested the median number of consecutively produced twitter syllables produced by the five siblings. We observed that the median number of syllables was significantly different between monkeys (P = 2.1 \times 10^{-11}, \chi^2 = 55.9, five animals, n = 1816, Kruskal-Wallis test), with a higher median syllable count for the S1 siblings than monkey S2-2 (P = 1.0 \times 10^{-8}, P = 0.014, and P = 0.037, post hoc multiple comparison test).

**Phee calls**

In contrast to trill and twitter calls, S1 animals could be clearly separated from the S2 litter by differences in the distributions of the median of all three measured parameters in most animals (Fig. 1E). The vocalizations of S1 monkeys were significantly shorter and had a lower frequency than calls from S2 monkeys, although monkey S1-3 did not show any significant difference from monkey S2-2 (duration: all S1 versus S2-1: P = 3.5 \times 10^{-73}, P = 4.9 \times 10^{-90}, and P = 8.7 \times 10^{-2}; all S1 versus S2-2: P = 4.8 \times 10^{-3}, P = 1.5 \times 10^{-15}, and P = 0.277; frequency: all S1 versus S2-1: P = 4.1 \times 10^{-85}, P = 9.2 \times 10^{-65}, and P = 2.5 \times 10^{-5}; all S1 versus S2-2, P = 1.3 \times 10^{-10}, P = 1.4 \times 10^{-7}, and P = 0.646; post hoc multiple comparison test). In contrast, all limited parental interaction monkeys (S1) exhibited phee calls with higher entropy values than both normally raised siblings (all S1 versus S2-1: P = 6.6 \times 10^{-17}, P = 3.7 \times 10^{-42}, and P = 8.3 \times 10^{-3}; all S1 versus S2-2: P = 0.012, P = 4.9 \times 10^{-16}, and P = 0.061; post hoc multiple comparison test).

As shown in previous studies, the peak frequency and Wiener entropy decrease throughout maturation (11, 14, 17). In contrast, call duration does not significantly change for trill and twitter calls (17), whereas it increases for phee calls during maturation (17). Previous studies of body size and call frequency allometry show a strong negative correlation between these two parameters in mammals, including marmoset monkeys (23, 24). Decreases in call frequency could be completely explained by the growth of the vocal tract (23, 25). Moreover, lower entropy values in all call types (11, 14) and longer durations specifically in phee calls (17) indicate vocal maturity in marmoset monkeys. The calls of the limited parental interaction monkeys showed more mature acoustic properties in trill and twitter calls compared with those of their normally raised siblings, which can be directly related to differences in age and body size.

However, for phee vocalizations, the distributions of call entropy and duration were of a less mature pattern in the older monkeys with limited parental interaction when compared with their normally raised siblings. Recent studies indicate that call entropy in particular is an ideal proxy to measure maturation during vocal development. Higher Wiener entropy corresponds to more broadband and therefore less mature calls, whereas lower entropy indicates more narrowband, tonal, and mature calls (11, 12, 14).

**Comparison of babbling-like sequences and individual calls in S1 monkeys**

In a previous study, we observed infant-specific sequences, babbling behavior as it is most commonly referred to in the literature (15–19), in monkeys with limited parental interaction (12). This behavior, which usually disappears after the first couple of months of vocal development (11, 17), is defined by vocal sequences with mature and immature call types, which are consecutively produced with intercall intervals in between 100 and 500 ms (Fig. 2A) (15, 20). Although marmoset babbling behavior shows a number of similarities to human babbling behavior with respect to universality, repetitiveness, rhythmicity, and the use of a subset of the adult vocal repertoire (26), it is important to note here that there are fundamental differences between both behaviors. For example, although human babbling represents some aspects of vocal
maturity, that is, by producing speech-like consonant-vowel sounds, marmoset babbling is a mixture of immature and mature vocalizations (19). Therefore, we refer to the observed infant-specific sequences as babbling-like behavior.

In addition to such babbling-like sequences, monkeys with limited parental interaction produce individual vocalizations just as normally raised monkeys do with intercall intervals that are longer than 500 ms (Fig. 2A). These observations are reflected in the distribution of intercall intervals in S1 and S2 monkeys. As previously shown (12), intercall intervals were significantly different between monkeys ($P = 1.0 \times 10^{-227}$, $\chi^2 = 1057.85$, five animals, 11,370 vocalizations, Kruskal-Wallis test). However, post hoc cluster analysis clustered S1 and S2 siblings in their respective litter (weighted linkage, proximity matrix with Spearman distance). The main difference observed between S1 and S2 monkeys was the lack of intercall intervals between 100 and 500 ms, which is due to the lack of babbling-like behavior in S2 monkeys (Fig. 2B).

As a next step, we evaluated differences in call entropy between the S1 monkeys and between babbling-like sequences and individual vocalizations. We excluded monkey S1-3 from data analyses because of a small number of individually uttered vocalizations and the lack of babbling-like behavior in S2 monkeys (Fig. 2B). As we have shown, the Wiener entropy of phee calls produced by monkeys with limited parental interaction is significantly higher than that produced by S2 animals. Furthermore, phee entropy is higher during babbling-like behavior compared with individually uttered phees in S1 monkeys, suggesting that phee calls in particular undergo distinct and discrete changes throughout vocal development. A recent study suggested that adult phee calls are a transformation of infant-specific cries and subharmonic phees (14), which all coexist to a certain degree during the subadult stages (Fig. 3A) (11, 15). According to the model, parental feedback accelerates the maturation process during the cry-phee transition by decreasing call entropy. Our limited parental interaction animals produce all three call types simultaneously (Fig. 3A). We therefore have the unique ability to compare the call entropy of all maturation stages within these animals and to evaluate differences in phee call production in S2 monkeys.

Both S1 monkeys that produced a sufficient number of individual vocalizations showed significant differences in call entropy distributions between call types (cries, subharmonic phees, phees produced during babbling-like behavior, and individual phees; $P = 2.1 \times 10^{-252}$, $n = 992$ vocalizations, two-way ANOVA; Fig. 3B). Distributions of call entropy showed a stepwise decrease in call entropy with the highest entropy measures for cries followed by subharmonic phees and phees produced during babbling-like behavior, with individual phees showing the lowest entropy values in both monkeys. Finally, we tested for differences in the call entropy distribution of individual phees between the four monkeys.

---

**Fig. 2. Differences in vocal sequence production between S1 and S2 monkeys.** (A) From top to bottom: Spectrograms of an exemplar babbling-like sequence of an S1 monkey and exemplar vocal sequences with individual calls of an S1 and an S2 monkey, respectively. White horizontal lines indicate single calls and multisylilabic calls, respectively. (B) Relative distributions of intercall intervals (ICIs) from monkey S1-1 (red), S1-2 (orange), and S1-3 (pink), as well as S2-1 (dark blue) and S2-2 (light blue) (sliding window size, 100 ms; step size, 10 ms). (C) Median entropy distributions of trill, twitter, and phee calls made by S1-1 and S1-2 individually or during babbling-like behavior. Horizontal lines inside the boxes represent the median. Boxes represent the interquartile range (25th to 75th percentile), and whiskers indicate the 3rd and 97th percentiles.
Finally, we wanted to know how phee calls and their immature versions, babbling-like behavior motif-like sequences during marmoset trills being uttered the most, followed by tsiks, and infant cries being uttered almost three times as much by monkey S1-1 and more than twice as much by monkey S1-2 as would be expected from the overall call type distribution within babbling-like behavior. Furthermore, these sequences were most likely preceded by trill calls, occurring 1.5 times more than this trill-tsik-cry combination would be statistically expected for both monkeys (S1-1: $P = 1.3 \times 10^{-9}$, $\chi^2 = 44.3$; S1-2: $P = 1.6 \times 10^{-8}$, $\chi^2 = 39.1$; $\chi^2$ test; Fig. 4C). Finally, call distributions preceding these three-call sequences were also significantly different from what would be expected from the overall distribution. In both monkeys, the occurrence of a tsik was 50% higher than expected. In addition, a 50% higher probability for cries was already noted in one of the siblings (S1-1: $P = 5.5 \times 10^{-7}$, $\chi^2 = 31.9$; S1-2: $P = 8.9 \times 10^{-6}$, $\chi^2 = 26.2$; $\chi^2$ test; Fig. 4C). This motif-like ABC and BABC structure was quite stable over several repetitions. It could be observed regularly within the babbling-like behavior and was sometimes observed to be extended to longer repeated syllables (Fig. 4D, lower trace). Occasionally, we observed that cries were replaced by subharmonics or phees within this motif-like structure (Fig. 4E), suggesting an interchangeable use of all developmental transitions of phees, such as cries, subharmonic phees, and adult phees, within the babbling-like behavior of our subjects.

DISCUSSION

We show that limiting direct parental interaction has long-term effects on the acoustic properties of vocalizations in marmoset monkeys. Although all monkeys produced mature call types, the vocal utterances of marmoset monkeys, which experienced limited parental interaction during infancy, were characterized by phee vocalizations at a subadult stage that showed acoustic immaturity in distinct call parameters in comparison with the phee calls of normally raised monkeys. These findings suggest that limiting direct parental interaction has long-term effects on vocal production–related call structure features and indicate a significant role of social interaction on the proper development of the acoustic structure of marmoset vocalizations.

The phee calls of all monkeys that experienced limited parental interaction exhibited significantly higher entropy values than their younger, normally raised siblings. In addition, the calls were also significantly shorter. In contrast, the call entropy and duration of the other two investigated call types did not show any systematic differences between S1 and S2 monkeys. Call frequencies were significantly lower in S1 animals for all three call types. In our experiments, siblings with limited social interaction were slightly older and heavier than their normally raised siblings. The observed lower call frequencies in S1 monkeys were therefore attributed to their heavier weight, because various studies have shown a strong negative correlation between body size/weight and call frequency (body size–frequency allometry) in several mammals, including marmoset monkeys (14, 23–25). Similarly, the higher number of syllables in twitter calls of the older S1 monkeys can be directly correlated with age, because a higher median number of trill syllables were exhibited by the older and, therefore, heavier subadult marmosets (17). Similar to call frequency, lower entropy values are associated with vocal maturity (3, 14) and therefore characterize the vocal utterances of older (and heavier) monkeys. However, call entropies in the phee calls of S1 monkeys were still higher than those of S2 animals, indicating less maturity in the older S1 monkeys. These differences in the Wiener entropy of phee calls could

Fig. 3. Transformation of infant marmoset calls into adult phee calls. (A) Examples of a cry, subharmonic phee, and proper phee call. As recently suggested, infant-specific cries transform to subharmonics and phees, and subharmonics transform to phees during vocal development and coexist to a certain degree during the subadult stages. (B) Median entropy distributions of distinct call types of S1 monkeys indicating different levels of maturity (cries, subharmonic phees, phees produced in babbling-like behavior, and individual phees) and proper phee calls of S2 monkeys. Horizontal lines inside the boxes represent the median entropy of cries, subharmonic phees, phees produced in babbling-like behavior, and individual phees of the S1-1 (red) and S1-2 (orange) subjects and the phees of the S2-1 (blue) and S2-2 subjects (light blue). Boxes represent the interquartile range (25th to 75th percentile), and whiskers indicate the 3rd and 97th percentiles. (S1-1, S1-2, S2-1, and S2-2) to evaluate potential differences in maturational levels between S1 and S2 animals. We observed significant differences in the call entropy distribution of individual phee vocalizations between monkeys ($P = 2.1 \times 10^{-17}$, $\chi^2 = 80.8$, four animals, $n = 727$ vocalizations, Kruskal-Wallis test). Post hoc analyses revealed that S1-2 exhibited significantly higher call entropy values than the S2 monkeys ($P = 1.6 \times 10^{-8}$ and $P = 3.4 \times 10^{-3}$, post hoc multiple comparison test). The S1-1 monkey showed higher call entropies with higher variance than its S2-1 sibling; however, the differences were not significant ($P > 0.5$, post hoc multiple comparison test).

Motif-like sequences during marmoset babbling-like behavior

Finally, we wanted to know how phee calls and their immature versions, cries/compound cries and subharmonic phees, are used within the babbling-like sequences of S1 monkeys. If the recently proposed model (11, 13) is correct, then these call types should be produced interchangeably. Furthermore, if specific repetitive and predictable patterns exist within babbling-like behavior, then we should be able to evaluate this. Babbling-like behavior consists of bouts composed of long sequences of a rhythmically uttered mixture of mature call types, such as trills and tsiks, and immature calls, such as infant cries (Fig. 4A). Here, both animals show similar distributions of the investigated call types, with trills being uttered the most, followed by tsiks, and infant cries being elicited at least often (S1-1: $P = 0$, $\chi^2 = 1455.1$; S1-2: $P = 8.7 \times 10^{-172}$, $\chi^2 = 794.0$; $\chi^2$ test; Fig. 4B).

We evaluated whether these call types were uttered randomly or whether specific sequences were more common than others. For this, we compared the relative occurrence of distinct call transitions as a function of how often we would expect these transitions based on the relative distribution of call types within babbling-like sequences. We observed that, for both animals (S1-1 and S1-2), infant cries were predominantly preceded by tsik calls (S1-1: $P = 2.3 \times 10^{-48}$, $\chi^2 = 224.4$; S1-2: $P = 1.6 \times 10^{-48}$, $\chi^2 = 206.6$; $\chi^2$ test; Fig. 4C). These tsik-cry sequences were uttered almost three times as much by monkey S1-1 and more than twice as much by monkey S1-2 as would be expected from the overall call type distribution within babbling-like behavior. Furthermore, these sequences were most likely preceded by trill calls, occurring 1.5 times more than this trill-tsik-cry combination would be statistically expected for both monkeys (S1-1: $P = 1.3 \times 10^{-9}$, $\chi^2 = 44.3$; S1-2: $P = 1.6 \times 10^{-8}$, $\chi^2 = 39.1$; $\chi^2$ test; Fig. 4C). Finally, call distributions preceding these three-call sequences were also significantly different from what would be expected from the overall distribution. In both monkeys, the occurrence of a tsik was 50% higher than expected. In addition, a 50% higher probability for cries was already noted in one of the siblings (S1-1: $P = 5.5 \times 10^{-7}$, $\chi^2 = 31.9$; S1-2: $P = 8.9 \times 10^{-6}$, $\chi^2 = 26.2$; $\chi^2$ test; Fig. 4C). This motif-like ABC and BABC structure was quite stable over several repetitions. It could be observed regularly within the babbling-like behavior and was sometimes observed to be extended to longer repeated syllables (Fig. 4D, lower trace). Occasionally, we observed that cries were replaced by subharmonics or phees within this motif-like structure (Fig. 4E), suggesting an interchangeable use of all developmental transitions of phees, such as cries, subharmonic phees, and adult phees, within the babbling-like behavior of our subjects.

Gultekin and Hage, Sci. Adv. 2018;4:eaar4012 | 11 April 2018
not be simply explained by different states of arousal in the two different types of vocal behavior, that is, babbling-like behavior and individual calling, because distributions of call entropy did not show any (trill and phee) or systematic differences (twitter) when uttered during babbling-like behavior or individual calling, respectively. The ability to produce lower frequency values indicates that all S1 and S2 monkeys were physically mature per se, which is also supported by their ability to produce adult call types. Therefore, it is unlikely that the immaturity in...
some vocal production–related acoustic call features, such as duration and entropy, is related to physical, internal factors in the S1 monkeys.

It is generally established that within the framework of vocal communication systems, there are three types of vocal learning: comprehension, usage, and production (22, 27, 28). Auditory comprehension learning is characterized by the ability to associate a distinct auditory cue with an adequate behavioral response or objects in the environment and is broadly distributed among vertebrates, including primates (22, 28). Vocal usage learning is defined by the ability to volitionally control when and where, but not how, to produce a specific vocalization in a specific cognitive, social, or environmental context (28). Conceptually, one might distinguish two types of vocal usage learning. The first is the ability to withhold or initiate a specific vocalization, although it is still tied to the respective (motivational) context (29, 30). The second is the more elaborate ability to decouple calls of the monkeys’ vocal repertoire from the accompanying motivational state for use in a novel context (31). For example, infant marmosets produce some call types in inappropriate contexts (32) and seem to learn the appropriate context with experience during vocal development (12). Another example is rhesus monkeys that are capable of instrumentalizing their vocal output to successfully perform a specific task in operant conditioning tasks (33, 34). Vocal production learning is characterized by experience-dependent changes of the acoustic structure of vocalizations (22), for example, the ability to generate new vocal patterns, and it is crucial for human speech development. Besides humans, vocal production learning has been widely found in songbirds, parrots, cetaceans, bats, and a few other lineages (22, 35). The existence of vocal production learning mechanisms in nonhuman primates has been a matter of debate for decades (36, 37). Vocalizations of nonhuman primates have been thought for decades to be largely innate and fixed at birth and therefore believed not to undergo production–related acoustic changes that may be related to external factors (8, 37). In this regard, Seyfarth and Cheney (38) noted that 79% of the studies that claim that the call structure of nonhuman primates is fixed at birth were published before 1987; after 1987, 71% of the studies published reported some sort of developmental modification in vocal development. Moreover, call entropy, which has already been used for the characterization of distinct bird song features for almost two decades (3), is curiously a relatively novel acoustic feature to analyze in the vocalizations of nonhuman primates (11, 14) and decreases throughout vocal development, indicating maturity in vocal structure. Most earlier studies focused on frequency and duration, two parameters that are highly dependent on physical factors such as the length of the vocal tract, body size, and lung volume (23–25). Thus, the long-held idea of nonhuman vocalizations being fixed at birth may have been established because of the acoustic parameters chosen for investigation and the lack of appropriate quantitative analytic methods and devices.

The vocalizations of marmoset monkeys undergo marked changes during maturation. Several call parameters, such as call duration, frequency, and entropy, undergo call type–specific changes during vocal development (11, 14, 17). Trill and twitter calls, two common call types, show a drop in peak frequency and entropy measures with maturation but do not show any major changes in call duration (17). However, it has been a matter of debate for decades whether these changes were purely a consequence of physical maturation or whether environmental factors have a direct impact on acoustic call structure during vocal development, as in humans and songbirds (8, 39). Recent studies have thoroughly investigated the vocal development of marmoset monkeys in the first postnatal weeks (11, 14, 20) and beyond infancy (12), showing that the vocal development of marmoset monkeys cannot be explained solely by physical maturation (11, 13). Recently, it was shown that infant marmosets with limited parental interaction exhibit immature vocal behavior (so-called infant-specific babbling-like behavior) beyond infancy, whereas their normally raised siblings from another litter exclusively produced the adult vocal repertoire (12). In addition, the proportion of the exhibited call types differed between both litters, indicating immaturity in vocal behavior in the limited parental interaction subjects (12). Another recent study showed that the transition from immature to mature vocal behavior is accelerated by experimentally increasing contingent parental feedback in marmosets, indicating experience-dependent changes of acoustic call structures (14). However, differences in acoustic structure that were exhibited by different amounts of contingent feedback disappeared after 35 postnatal days in the aforementioned study. Although all of these studies demonstrated that social parental interaction is essential for normal vocal development, all monkeys within these studies were ultimately able to produce adult call types, regardless of whether they showed faster or slower transitions to mature-like vocalizations. Our present results suggest a crucial role for social feedback in vocal production learning and provide experimental evidence that limiting parental feedback leads to long-term effects on the acoustic call structure in marmoset monkeys.

Furthermore, we show that a separation of marmoset infants from their caregivers after the third postnatal month severely affects the vocal behavior of the offspring. These results point to a potential experience-dependent extension of the “critical period” in marmoset vocal development, similar to the ones observed in humans and songbirds (40, 41). Earlier studies have shown that some acoustic call parameters follow changes in the social environment (42), suggesting that the observed vocal changes in monkeys S1-1 and S1-2 might be a result of the direct interaction of these monkeys with the hand-raised monkey S1-3. However, this potential explanation seems to be unlikely because vocal imitation or complex volitional modifications in the acoustic call structure have not yet been observed in any monkey species (8, 22, 31, 35, 36). In such a case, we would also expect the vocal parameters of the imitating monkeys to be between the values of normally raised S2 monkeys and the hand-raised monkey S1-3, which we did not observe in the present study. Future studies will have to elucidate how far this critical period reaches into the subadult stages of the marmoset life span and how inter-individual infant interaction modulates vocal output within this period.

In addition, our study supports a recently proposed model, suggesting a transformation from high-entropy, immature cry vocalizations to low-entropy, mature phee vocalizations (11). During babbling-like behavior, call types were not uttered in a random order but in motif-like sequences similar to motifs produced within bird song (4). These motif-like sequences were characterized by a higher occurrence of a specific call succession, “BABC” (tsik-trill-tsik-cry), which was uttered with a higher probability than any other call combination. Here, we have shown that cries were occasionally exchanged by trills or subharmonic phees within these motif-like sequences, indicating a similar usage of these call types within babbling-like bouts. This implies that cries and subharmonics produced during infant babbling-like sequences might be the immature versions of the adult phee calls. Our data show stereotypical, motif-like sequences in the babbling bouts of our subadult monkeys with limited parental interaction. Further studies will now have to decipher whether these stereotypical sequences are a universal vocal pattern, which can be observed in all infant marmosets, independent on whether they were raised normally or with limited...
parental feedback, or whether it is rather a result of a prolonged state of vocal immaturity in limited parental feedback monkeys.

However, limited parental vocal interaction might not be the only reason for these changes in the vocal behavior of young marmosets separated from their parents at an early age. Confounding aspects, such as environmental and psychological factors, might have caused the observed vocal abnormalities in these animals. First of all, separation of the subadult siblings from the parents was initially motivated by a chance to establish a stable social group with the hand-raised monkey instead of housing it alone. We attempted to minimize environmental differences by moving the S1 siblings into a cage with a largely identical layout within the same room as the parental cage. Furthermore, we paid particular attention to find the best separation time to reunite the S1 siblings. On the one hand, we wanted to ensure that the siblings stayed with their parents long enough to minimize potential psychological confounds, such as an elongation of immaturity (with an accompanied continuation of infant vocalizations). On the other hand, to increase the likelihood of successful reunification, it was important to reunite the three siblings at an early age. For the following reasons, we decided to reunite the siblings after the third month. After this period, weaning is largely completed and young marmosets spend most of their time away from their parents and locomote independently (43, 44). These findings were supported by our personal observations during the last 2 weeks before the separation, showing that the siblings were independently moving around and were barely being carried by their parents. In addition, they were able to independently eat out of food dishes and autonomously hunt small offered prey (for example, mealworms and locusts). We observed no aggressive behavior immediately after reunification. Instead, we witnessed affiliative interactions, such as grooming behavior, between the S1 siblings. This continued throughout the following months, indicating a healthy and stable social group (45). Abnormal or atypical behavior, other than the reported vocal differences, has not been observed. In this context, it is important to note that the hand-raised monkey and the two siblings spent different amounts of time with their parents before the separation and thus might have had different psychological experiences during this time. However, the vocal behavior was more similar within the S1 litter than it was between the two normally raised monkeys (12). Therefore, we assume that psychological and environmental factors had only minor effects on the vocal behavior of the S1 siblings and that the observed vocal differences between groups were correlated to differences in direct parental interactions rather than other factors. Finally, we collected a large number of vocalizations to obtain a reliable data set for statistical analyses, thus compensating for the relatively small sample size. In addition, the control group litter was from the same parents as the S1 subjects. This approach provided an opportunity to compare monkeys with equal genetic relatedness and to largely exclude genetic variation as a main confounding factor for differences in vocal behavior between litters.

In humans, social interactions play a crucial role in the acquisition of speech sounds. Changes in the acoustic structure (1) and lack of social interaction lead to severe limitations in speech acquisition (2). Visual and acoustic information provided during social interactions between caregivers and infants enhances the acquisition of speech sounds, whereas passive exposure to speech alone leads only to minimal vocal learning (1, 2). Similarly, vocal learning is also highly influenced by visual and auditory information during social interactions in songbirds (3, 4, 46). However, despite the essential role of social interaction on vocal learning, little is known about the underlying neural processes.

Our study provides new evidence that limiting parental interaction after the first postnatal months leads to long-term alterations in the acoustic structure of distinct call types, indicating the presence of vocal production learning in marmoset monkeys. Future studies will now have to decipher the neural mechanisms regulating the maturation of phee calls and the influence of direct parental feedback. Their highly vocal nature and rare cooperative breeding system make the marmoset monkey a compelling model system to investigate the neurophysiological principles underlying early vocal development in humans.

MATERIALS AND METHODS

Experimental design

Experimental animals

Here, we used five captive marmoset monkeys (Callithrix jacchus) born in two different litters to the same parents that have already been used in our previous study (12): one set of 13-month-old subadult triplets (S1) and one set of 7-month-old subadult twins (S2). As marmosets typically give birth to dizygotic twins (47), siblings from different litters are genetically as similar as same-litter siblings. The male monkey of the triplets (S1-1) was disowned by the parents on the third day and hand-raised by an animal caretaker for the first 3 months. During this time, the monkey was placed in a small cage attached to the parents’ home cage, allowing visual and auditory parental feedback for approximately 6 to 8 hours per day. The other two monkeys from this set (both females, S1-2 and S1-3) were raised by their parents until the end of the third month. After 3 months, we separated the two females and reunited them with their hand-raised sibling in a separate cage within the animal facility. In contrast, the second set of marmosets [twins: S2-1 (male) and S2-2 (female)] remained grouped with their parents during the vocal recording period.

The marmoset monkeys in our colony were all born in captivity and are held in pairs or family groups. The colony room has a 12-hour day/12-hour night cycle with a temperature of approximately 27°C and 55 to 65% relative humidity. The animals had ad libitum access to water and were fed daily with commercial pellets, curd cheese, fruit, vegetables, mealworms, and locusts. Additional treats, such as marshmallow juice or fruit juice, were used as positive reinforcements to transfer the animals from their home cage to the experimental cage. All procedures were authorized by the national authority, the Regierungspräsidium Tübingen, Germany.

Vocal recordings

The vocal recordings used in this study are a fraction of the vocalizations that were reported in a previous paper (12). We recorded the vocal behavior of subadult marmosets when they were separated from their social group. For this purpose, animals were trained to enter an experimental cage (25 cm × 25 cm × 28 cm), in which they were separated from their social group. During vocal recordings, animals, which were raised by their parents for at least 3 months (S1-2, S1-3, S2-1, and S2-2), had visual and acoustic contact with a pair of unrelated adult marmoset monkeys that were placed in a cage at approximately 1.5 to 2 m. The hand-raised monkey (S1-1) produced infant vocalizations toward its main caregiver during infancy, an animal caretaker, but did not do so toward other marmoset monkeys. We never observed the production of infant-specific vocal sequences while the S1 monkeys were in their home cage interacting with each other. Pygmy marmoset infants produce infant vocalizations, especially in infant-caregiver interactions, because the caregivers are more likely to approach them when they do so (26). Therefore, we recorded monkey S1-1 while having visual and
acoustic contact with the animal caretaker. Vocalizations were recorded using a condenser microphone (Sennheiser MKH 8020 with pre-amplifier MZX 8000), which was placed 10 cm in front of the small cage and digitized at a sampling frequency of 96 kHz via an A/D interface (Roland UA-1010 OctaCapture). Daily sessions lasted 10 to 15 min each.

Acoustic analysis
We analyzed the adult call types (“phee,” “trill,” and “twitter”) that were recorded in five daily sessions per individual monkey (5 monkeys, 25 sessions, and 8497 vocalizations). As in our previous study (12), we did not aim to classify calls within one call type into subtypes. We classified marmoset vocalizations into already-defined main groups (12, 17). Calls were manually classified on the basis of their spectrotemporal profile and auditory playback. A twitter is a short upward frequency modulation (FM) sweep, which is normally produced in a multisyllabic matter. Trill calls are defined by a sinusoidal-like FM throughout the call. Phees are tone-like long calls with a fundamental frequency (F0) around 7 to 10 kHz. Trill-phees, which are the combination of two call types, trill and phee, were classified into the call type that was predominantly present within the call. A cry is a broadband call with an F0 of around 600 Hz. A compound cry is a combination of a cry and another call, irrespective of the order. Cries and compound cries are very similar in call structure. We have therefore referred to cries and compound cries as proper cries throughout the manuscript. A subharmonic phee is similar to a phee but has a visible subharmonic component (first subharmonic around 3.5 to 5 kHz). A tsik is a broadband short call consisting of a linearly ascending FM sweep that merges directly into a sharply descending linear FM sweep.

Call onsets and offsets were manually detected using Avisoft-SASLab Pro 5.2 software (Avisoft Bioacoustics). Duration, peak frequency, and Wiener entropy values were extracted using the same software. Duration was calculated as the time between the beginning and the end of a vocalization. The intercall interval was defined as the time between the end of a vocalization and the beginning of the consecutive utterance. The peak frequency of a call was defined as the frequency with the maximum amplitude within the spectrum. The Wiener entropy was used as a measure for how broadband the power spectrum of a specific sound is (corresponding to the noisiness of the sound) and was calculated as the logarithm of the ratio between the geometric and arithmetic means of the values of the power spectrum (3, 11, 48). Noisy sounds exhibit high Wiener entropy values, whereas tonal sounds exhibit low entropy values. The signal was band-pass–filtered between 5 and 15 kHz, because the vast majority of marmoset calls fall into this range (16). For analyses, the spectrograms were calculated using a fast Fourier transform window of 1024 points, a Cepstrum (15, 20), with intercall intervals of more than 500 ms were flagged as “individual vocalizations.” Intercall intervals of less than 100 ms correspond to the intervals between syllables produced within multisyllabic vocalizations such as twitter calls (20). Here, syllables of multisyllabic vocalizations were classified as vocalizations uttered during babbling-like behavior or “individual” calls dependent on the time between the first and the last syllable of the call to the preceding and consecutive vocalizations, respectively.

Statistical analyses
We performed Kruskal-Wallis tests to reveal differences in duration, peak frequency, and entropy values between call types from different subjects. When significant differences occurred, we tested post hoc for differences between individual monkeys with a post hoc multiple comparison test (with the Bonferroni method). To test for differences in intercall intervals, we performed Kruskal-Wallis tests with a post hoc clustering analyses using weighted linkage (proximity matrix with Spearman distance) (49). Here, we focused on the range between 0 and 2 s (11,370 calls), because changes in intercall intervals were most prominent below 2 s. We used two-way ANOVAs to reveal differences in entropy values between the vocalizations produced during babbling-like behavior of the S1 siblings and individually, and we tested post hoc with multiple comparison tests to determine differences between individual monkeys. To test for significant differences between observed and expected call transition distributions, we performed χ² tests. In all performed tests, significance was tested at α = 0.05. When using the same data set multiple times (Kruskal-Wallis tests to reveal differences in duration, peak frequency, and entropy values of the same calls), we used Bonferroni correction, resulting in a corrected α = 0.017. Statistical analysis was performed using MATLAB (MathWorks).

REFERENCES AND NOTES
7. S. Derégnaucourt, C. Poirier, A. Van Der Kant, A. Van Der Linden, M. Gahr, Comparisons of different methods to train a young zebra finch (Taoiopsiodio guttata) to learn a song. J. Physiol. Paris 107, 210–218 (2013).

SUPPLEMENTARY MATERIALS
Supplemental material for this article is available at http://advances.sciencemag.org/cgi/content/full/4/4/eaar4012/DC1
fig. S1. Difference in number of twitter syllables per twitter call in normally raised monkeys and siblings with limited parental interactions.


Acknowledgments: We thank J. Holmes for proofreading. *Funding:* This work was supported by the Werner Reichardt Centre for Integrative Neuroscience (CIN) at the Eberhard Karls University of Tübingen (CIN is an Excellence Cluster funded by the Deutsche Forschungsgemeinschaft within the framework of the Excellence Initiative EXC 307). *Author contributions:* Y.B.G. conducted the experiments; both authors designed the experiments, analyzed and interpreted the data, and wrote the manuscript. *Competing interests:* The authors declare that they have no competing interests. *Data and materials availability:* All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 7 November 2017 Accepted 21 February 2018 Published 11 April 2018 10.1126/sciadv.aar4012

Limiting parental interaction during vocal development affects acoustic call structure in marmoset monkeys
Yasemin B. Gultekin and Steffen R. Hage

Sci Adv 4 (4), eaar4012.
DOI: 10.1126/sciadv.aar4012