

RESEARCH ARTICLE

Developmental changes of cognitive vocal control in monkeys

Steffen R. Hage^{*‡}, Natalja Gavrilov and Andreas Nieder

ABSTRACT

The evolutionary origins of human language are obscured by the scarcity of essential linguistic characteristics in non-human primate communication systems. Volitional control of vocal utterances is one such indispensable feature of language. We investigated the ability of two monkeys to volitionally utter species-specific calls over many years. Both monkeys reliably vocalized on command during juvenile periods, but discontinued this controlled vocal behavior in adulthood. This emerging disability was confined to volitional vocal production, as the monkeys continued to vocalize spontaneously. In addition, they continued to use hand movements as instructed responses during adulthood. This greater vocal flexibility of monkeys early in ontogeny supports the neoteny hypothesis in human evolution. This suggests that linguistic capabilities were enabled via an expansion of the juvenile period during the development of humans.

KEY WORDS: Communication, Speech evolution, *Macaca mulatta*, Vocal utterance, Volitional call initiation, Vocal flexibility

INTRODUCTION

The human language faculty vastly outperforms primate vocal communication systems in scope and flexibility (Balter, 2010; Ghazanfar, 2008; Hammerschmidt and Fischer, 2008). This lack of essential linguistic characteristics in extant non-human primate communication systems hampers insights into the evolutionary origins of speech and language (Arnold and Zuberbühler, 2006; Seyfarth and Cheney, 2010). Volitional control of vocal utterances is deemed a critical, albeit insufficient, precursor for the development of a flexible communicative system (Balter, 2010; Ghazanfar, 2008; Ackermann et al., 2014). However, primate communication systems consist of stereotyped and innate calls that are almost exclusively uttered affectively (Ackermann et al., 2014; Deacon, 2010; Jürgens, 2002). Non-human primates lack the neural machinery that endows modern humans with outstanding cognitive abilities such as language. The ‘neoteny hypothesis of human evolution’ (Gould, 1977) posits the expansion of the childhood period with refined synaptic development in modern humans to facilitate larger and more powerful neural systems. Specifically, the prefrontal cortex, which is associated with the highest levels of cognition in addition to being the site of Broca’s language production area, experiences extraordinary long phases of developmental reorganization of neuronal circuits (Petanjek et al., 2011). Genes related to the development of the prefrontal cortex show excessive, neotenic expression in humans relative to chimpanzees and rhesus macaques (Somel et al., 2013).

The neoteny hypothesis suggests an exploitation of greater neural plasticity early in ontogeny to fostering the neural underpinnings of high-level communication systems like language (Carroll, 2003; Oller, 2000). Interestingly, primate vocalizations experience ontogenetic changes. In infant and juvenile simian monkeys, calls are more variable (Hammerschmidt et al., 2001; Pistorio et al., 2006; Takahashi et al., 2015) and vocal-related learning, such as call usage and comprehension, is facilitated (Seyfarth and Cheney, 1986, 2010). Therefore, infant and juvenile monkeys seem to have an advantage and can use vocal communication signals more flexibly.

Earlier studies revealed that monkeys and apes can be trained to vocalize in operant conditioning tasks (Sutton et al., 1973, 1974, 1985; Trachy et al., 1981; Coudé et al., 2011; Koda et al., 2007). We recently reported that two juvenile rhesus monkeys can be trained with effort to instrumentalize their calls as a conditioned response in a simple detection task (Hage et al., 2013). All but one study, including our own, that indicated the age of the monkeys and apes were performed with juvenile animals (Sutton et al., 1973, 1974; Trachy et al., 1981; Koda et al., 2007; Hage et al., 2013). Based on the neoteny hypothesis, we hypothesized that juvenile monkeys with a more plastic brain would be better suited for volitional call production than adult monkeys. We here present a longitudinal study based on data collected from two monkeys over several years, investigating potential developmental trends of vocal behavior from the juvenile to the adult period.

MATERIALS AND METHODS

Experimental animals

We used two male rhesus monkeys, *Macaca mulatta* (Zimmermann 1780), aged 4.8 and 4.9 years and weighing 4.2 and 4.5 kg at the beginning of this long-term study, and aged 9.5 and 9.7 years and weighing 8.6 and 9 kg, respectively, at the end. All procedures were authorized by the national authority, the Regierungspräsidium Tübingen, Germany.

Behavioral protocols

Both monkeys were first trained to perform a vocal response task (Fig. 1A), i.e. a visual go/no-go detection task using their vocalizations as a response (Hage et al., 2013; Hage and Nieder, 2013, 2015). Briefly, the monkeys were required to vocalize cued by arbitrary visual stimuli (red or blue squares) to receive a reward. Monkey T was trained to utter ‘coo’ vocalizations; monkey C was taught to emit ‘grunts’. The two colors appeared with equal probability ($P=0.5$) and had no significant influence on call probability (Wilcoxon signed rank test, $P>0.1$ for both monkeys). Trials began when the monkey initiated a ‘ready’ response by grasping a bar. Then, a visual cue, indicating the ‘no-go’ signal (‘pre-cue’; white square, diameter 0.5 deg of visual angle) appeared for a randomized time of 1–5 s (time epoch 1 of monkey C with times between 0.5 and 5 s). During this period, vocal output had to be withheld. Next, in 80% of the trials, the visual cue was changed to a colored ‘go’ signal (red or blue square; diameter 0.5 deg of

Animal Physiology, Institute of Neurobiology, University of Tübingen, Tübingen 72076, Germany.

^{*}Present address: Neurobiology of Vocal Communication, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Tübingen 72076, Germany.

[‡]Author for correspondence (steffen.hage@uni-tuebingen.de)

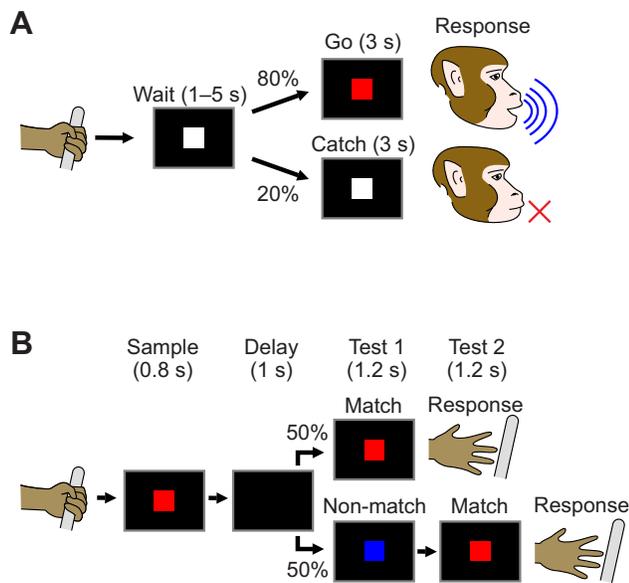


Fig. 1. Behavioral protocols. (A) In the vocal response task (visual go/no-go detection task), monkeys called within 3 s to indicate the detection of a color 'go' stimulus. They were required to withhold calls in the absence of a color go stimulus in catch trials. (B) In the manual response task (visual delayed match-to-sample task), monkeys released a bar (originally grabbed to initiate the trial) within 1.2 s to indicate the matching of a test color with the sample color. They were required to continue grasping the bar whenever a non-match color appeared in the test 1 period.

visual angle) lasting for 3000 ms (for monkey C, the duration of the go signal was extended to 3500 ms from the 19th session of epoch 6 until the end of epoch 8). During this time, the monkeys had to emit a vocalization to receive a reward. In 20% of the trials, the cue remained unchanged for another 3000 ms ('catch' trial). During this period, the monkey had to withhold calls. Catch trials were not rewarded. 'False alarms' were indicated by visual feedback (blue screen) and by trial abortion. To demonstrate its readiness to work, the monkey had to grab the bar throughout the pre-cue as well as the go phases. Bar release aborted the trials instantaneously, followed by visual feedback (red screen). In accordance with the go/no-go detection protocol, successful go trials were defined as 'hits', and unsuccessful catch trials as false alarms. One session was recorded per individual per day.

Vocal recording sessions comprised eight contiguous epochs in monkey C (epoch 1: median age 4.9 years with $N=15$ daily sessions; epoch 2: 5.4 years, $N=27$; epoch 3: 6.2 years, $N=25$; epoch 4: 6.8 years, $N=29$; epoch 5: 7.1 years, $N=47$; epoch 6: 7.7 years, $N=28$; epoch 7: 7.8 years, $N=13$; epoch 8: 8.0 years, $N=8$) and 7 epochs in monkey T (epoch 1: 4.9 years, $N=15$; epoch 2: 5.0 years, $N=33$; epoch 3: 5.8 years, $N=20$; epoch 4: 6.4 years, $N=52$; epoch 5: 6.7 years, $N=33$; epoch 6: 7.0 years, $N=12$; epoch 7: 7.8 years, $N=53$). Monkey T was head-fixed during all sessions, monkey C was head-fixed in all sessions during epoch 1–5. In both monkeys, epochs 2 and 5 include neuronal recording sessions, while all other sessions in the remaining epochs were behavioral sessions.

After the monkeys ceased to produce conditioned calls as a response, they were re-trained to perform a manual response task (Fig. 1B). They were trained to perform a standard visual delayed match-to-sample (DMS) task with colors and were required to respond to matching colors by hand movements. A trial started when the monkey grasped a lever. A sample display showing a color square (2 deg visual angle) was presented on a black background in

the center of a computer screen for 800 ms. A constant 1000 ms memory delay followed. Next, a test display appeared which in 50% of the cases was a match showing the same color as the sample period ('match' trials). In the other 50% of cases ('non-match' trials), the first test display after the delay period was a non-match, showing a different color, followed by a second test display, which always displayed a match color. If a match appeared, monkeys released the lever (within 1.2 s) to receive a fluid reward. If a non-match was shown, they held the lever until the second test display appeared (which in these trials was always a match), requiring a lever release for a reward. Trials were randomized and balanced across all relevant features (e.g. match versus non-match, colors). Monkey C performed the task with red and blue colors, monkey T with red, blue and green colors.

Data acquisition

As in our previous studies (Hage et al., 2013; Hage and Nieder, 2013, 2015), stimulus presentation and behavioral monitoring were automated on PCs running the CORTEX program (National Institutes of Health) and recorded by a Plexon Multi-Acquisition system. Vocalizations were recorded by the same system with a sampling rate of 40,000 Hz via an A/D converter. A custom-written MATLAB program running on another PC monitored the vocal behavior in real time and detected the vocalizations. Vocal onset times were detected offline by a custom-written MATLAB program to ensure precise timing for data analysis in all but two sessions of monkey C (epoch 3 and epoch 4), as these behavioral sessions were recorded by the CORTEX program only.

Recording of spontaneous vocalizations

The spontaneous vocalizations of the two monkeys in their housing environment were measured during their juvenile and adult periods as part of 'ethograms' for which a range of behaviors was recorded (Hage et al., 2014). To that aim, we focally sampled the call behavior of the monkeys in 1 min intervals over a duration of 10 min, during two periods of five consecutive days ('continuous sampling'; Altmann, 1974; Martin and Bateson, 1993). Call occurrence (%) could range from 0% (no calls during the 10 min observation window) to 100% (calls every minute during the 10 min observation) and was averaged for the juvenile and adult test periods. The data for the juvenile phase were collected when monkey C was 5.4 and 5.6 years old, and when monkey T was 5.0 and 6.1 years old. Spontaneous call behavior for the adult phase was recorded when monkey C was 9.5 years old, and when monkey T was 9.7 years old. Wilcoxon rank sum tests were used to test for significant differences in spontaneous vocal behavior between the juvenile phase and adulthood.

Statistical analysis

We computed d' -prime (d') sensitivity values derived from signal detection theory (Green and Swets, 1966) by subtracting z -scores (normal deviates) of median hit rates from z -scores of median false alarm rates. The detection threshold for d' values was set to 1.8. The d' criterion for the threshold was 1.8, which corresponds to a hit rate of 56% at a false alarm rate of 5% in this go/no-go task (Green and Swets, 1966).

Kruskal–Wallis tests (with *post hoc* Wilcoxon rank sum tests) were performed to test for significant differences in call performance, hit rate, false alarm rate, d' value and call latency during the detection task over time. We used Pearson's correlations to test for possible correlations between these parameters characterizing vocal behavior and the monkeys' age in the appropriate sessions.

RESULTS

We measured vocal behavior over a period of about 4 years, when monkey C's age ranged from 4.8 to 8.1 years and monkey T's age spanned from 5.1 to 7.9 years. During this time, we recorded 12,769 vocalizations in monkey C and 21,029 vocalizations in monkey T, which were uttered as obligatory responses in the vocal response task (Fig. 1A). In total, this corresponded to 192 daily sessions in monkey C, and 218 sessions in monkey T. Vocal recording sessions comprised eight contiguous epochs in monkey C and 7 epochs in monkey T (see Materials and methods for details). Fig. 2 shows the vocalization behavior of both monkeys over this time in relation to the timing of life history events in macaques (Fleagle, 2013). We measured several behavioral parameters characterizing call behavior: the total number of volitional calls per session, the hit rate (percentage correct responses) and the false alarm rate (vocalizations during catch trials without a go stimulus). The hit rate and false alarm rate were used to calculate the sensitivity index, or d' , from signal detection theory (Green and Swets, 1966). During the first epoch, and at an age of 4.8 and 5.1 years for monkey C and monkey T, respectively, both monkeys showed superior vocalization behavior. This was evidenced by high call rates

(monkey C: median 90 calls per session, Fig. 2A; monkey T: median 181 calls per session, Fig. 2B), high hit rates (monkey C: median 62.7%, Fig. 2C; monkey T: median 56.2%, Fig. 2D) and no false alarms at all in both monkeys (Fig. 2E,F). As a result of this high performance, the d' value was 4.0 in monkey C (Fig. 2G) and 3.9 in monkey T (Fig. 2H), and thus well above chance.

However, vocal performance progressively declined with increasing age of the monkeys. The number of calls per session decreased systematically over the epochs until both monkeys stopped uttering vocalizations completely (Fig. 2A,B; Kruskal–Wallis test; monkey C: $P<0.001$, $N=192$, $d.f.=7$, $\chi^2=109.1$; monkey T: $P<0.001$, $N=218$, $d.f.=6$, $\chi^2=138.1$) and was significantly correlated with age (Pearson's correlation: monkey C: $P<0.001$, $N=192$, $R=-0.63$; monkey T: $P<0.001$, $N=192$, $R=-0.63$; Fig. 2A,B). A similar decline of hit rates was observed for monkey C (Fig. 2C; Kruskal–Wallis test, $P<0.001$, $N=192$, $d.f.=7$, $\chi^2=125.1$) and monkey T (Fig. 2D; Kruskal–Wallis test, $P<0.001$, $N=218$, $d.f.=6$, $\chi^2=151.4$), which was also significantly correlated with age in both monkeys (Pearson's correlation: monkey C: $P<0.001$, $N=192$, $R=-0.80$; monkey T: $P<0.001$, $N=218$, $R=-0.73$; Fig. 2C,D). Importantly, however, the false alarm rate stayed at low

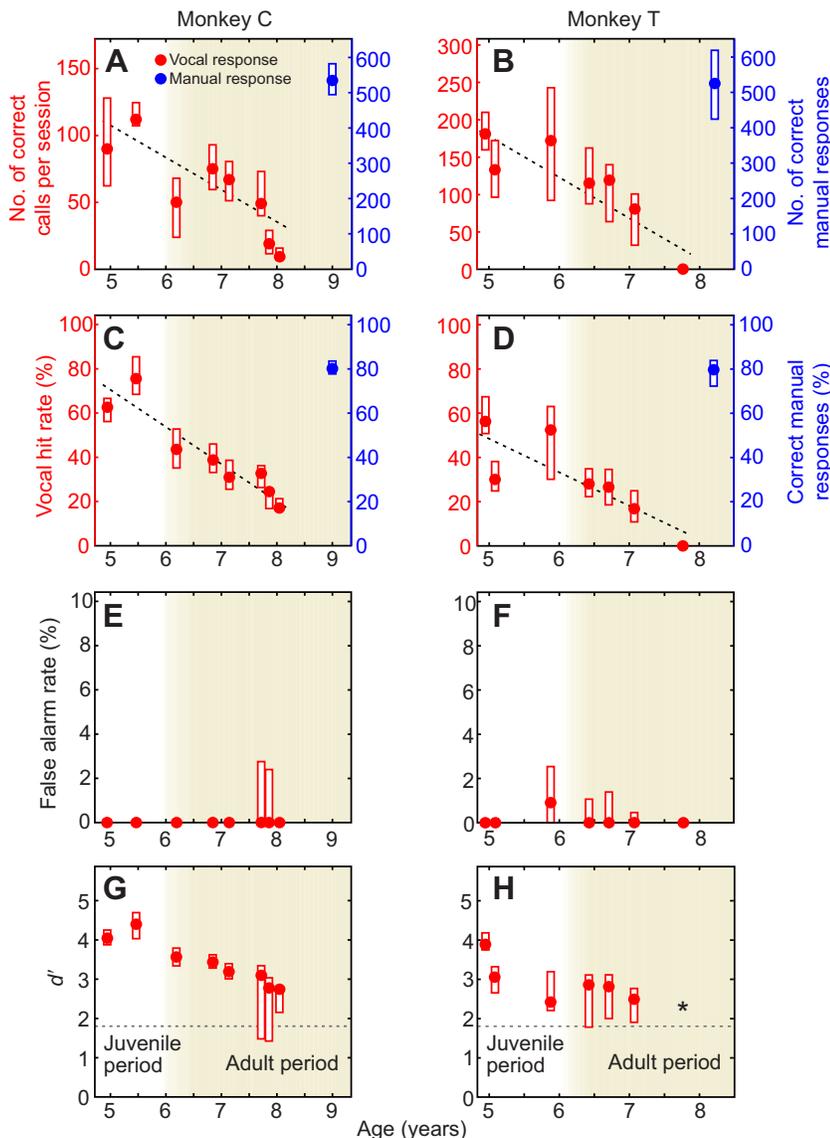


Fig. 2. Temporal trajectories of vocal behavior.

(A,B) Distribution of the number of vocalizations produced during the vocal response task (red bars) and number of performed trials in the manual response task (blue bars) of monkey C and monkey T within the recorded time epochs. (C,D) Distribution of hit rates within the vocal response task and correct responses in the manual response task. Dashed lines in A–D indicate significant correlations between call and hit rates and the monkey's age. (E,F) Distribution of false alarm rates in the vocal response task. (G,H) Distribution of d' values as a measure of sensitivity. Dotted lines indicate the d' threshold criterion of 1.8. *No d' value could be calculated for the last time epoch of monkey T because of the absence of vocal performance. Colored dots inside boxes indicate medians; lower and upper margins of boxes represent the first and third quartile, respectively. Note the different scales in A–D for the detection task (left) and delayed match-to-sample task (right). The shaded background indicates the adult period (Fleagle, 2013).

levels for all epochs in both monkeys (Fig. 2E,F), indicating that the monkeys did not develop arbitrary calling behavior. Therefore, the accompanying significant change of d' values (Fig. 2G,H; Kruskal–Wallis test; monkey C: $P<0.001$, $N=192$, $d.f.=7$, $\chi^2=118.3$; monkey T: $P<0.001$, $N=165$, $d.f.=5$, $\chi^2=42.6$), as well as the correlation of d' values with age, was caused by the decrease in overall vocalizations until extinction. However, median d' values were well above detection threshold until the end of the recordings (Pearson's correlation: monkey C: $P<0.001$, $N=190$, $R=-0.68$; monkey T: $P<0.001$, $N=165$, $R=-0.39$; Fig. 2G,H).

In parallel with the decline in performance, call latency increased significantly in duration. In monkey C, call latency changed from a median of 1.64 s in epoch 1 to 2.63 s in epoch 7 (Fig. 3A,B; Kruskal–Wallis test, $P<0.001$, $N=130$, $d.f.=4$, $\chi^2=91.0$, *post hoc* Wilcoxon rank sum test, $P<0.001$, $N=28$). In addition, median call latency was significantly correlated with the age of monkey C (Pearson's correlation, $P<0.001$, $N=130$, $R=0.77$). A less pronounced but equally significant decrease of call latency was observed between the first and last time epoch in monkey T (Fig. 3C; Kruskal–Wallis test, $P<0.001$, $N=165$, $d.f.=5$, $\chi^2=32.4$, *post hoc* Wilcoxon rank sum test, $P<0.02$, $N=68$). Changes in call latency did not constantly increase from epoch to epoch as in monkey C and showed only a weak, yet significant, correlation with the animal's age (Pearson's correlation, $P<0.01$, $N=165$, $R=0.21$).

To see whether the absence of vocalizations within the vocal response task was due to a general loss of vocal behavior, we investigated the spontaneous vocal behavior of both monkeys in their housing environment during their juvenile phase and adulthood. Fig. 4 depicts the mean occurrence of the monkeys' vocal behavior during focal animal scanning (10 min ethogram). Spontaneous calling behavior remained stable in monkey C (Wilcoxon rank sum test, $P>0.1$, $N=20$). Monkey T showed

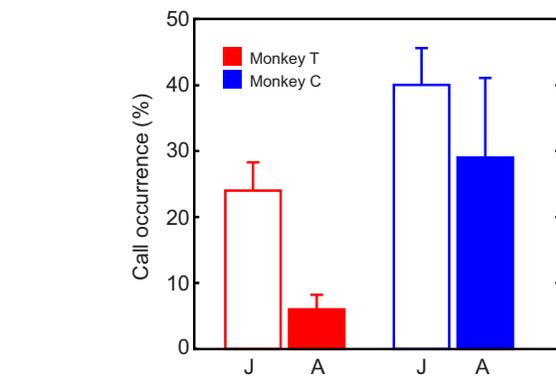


Fig. 4. Comparison of the occurrence of vocal behavior in the juvenile and adult phase for both monkeys. The bars show the mean+s.e.m. call behavior of each monkey within 10 min observation periods (100% indicates calls every minute during the 10 min; 20 sessions per monkey) as a function of the monkeys' developmental stage. J, juvenile; A, adult.

reduced spontaneous calling behavior during adulthood (Wilcoxon rank sum test, $P<0.01$, $N=20$), but never stopped vocalizing spontaneously. Thus, the ongoing spontaneous call behavior of both monkeys was in stark contrast to the complete halt of volitional vocalizations with age. Therefore, the reported decline of volitional vocalizations cannot be explained by a general lapse of calling behavior, because the monkeys continued to vocalize spontaneously in their housing environment, i.e. outside of the behavioral protocol.

Moreover, the discontinuation of volitional vocal behavior could also not be accounted for by major environmental changes. Throughout these years of training, the monkeys maintained continuous good health (also verified by regular blood tests) and gained normal weight. Moreover, the same behavioral protocol was presented, the same controlled fluid intake protocol for motivation was applied, the same housing of the monkeys in small social groups was carried out, and the same scientific trainers (S.R.H. and N.G.) worked with the monkeys throughout this 4 year period.

Finally, we wondered whether the extinction of volitional calling could be explained by a general loss of volitional responses, a lack of motivation, or some general resistance to respond in a conditioned task. To test this possibility, we re-trained both monkeys after they stopped vocalizing in the vocal response task on a manual response task. To remain within the same sensory modality, we trained them to perform a DMS task with color stimuli (Fig. 1B). Monkeys were required to use a manual bar release instead of a vocalization as a response. Even though a DMS discrimination task is more demanding in comparison to the previous simple detection task, the monkeys, which were now 9.0 years old (monkey C) and 8.2 years old (monkey T), showed full recovery of the volitional response. Monkey C performed, on average, 534 trials (8 sessions) and monkey T performed 526 trials (7 sessions; both medians; Fig. 2A,B). They also showed a high median percentage of correct responses (Fig. 2C,D; monkey C: 80.2%, monkey T: 79.7%). Both monkeys continued to work at this high performance level.

DISCUSSION

We report a systematic decline of volitional vocalizations in rhesus monkeys that was not explained by (a) a general lapse of calling behavior, (b) environmental changes or (c) a general loss of voluntary responses or lack of motivation. During this longitudinal investigation, we also performed unilateral single-unit recordings

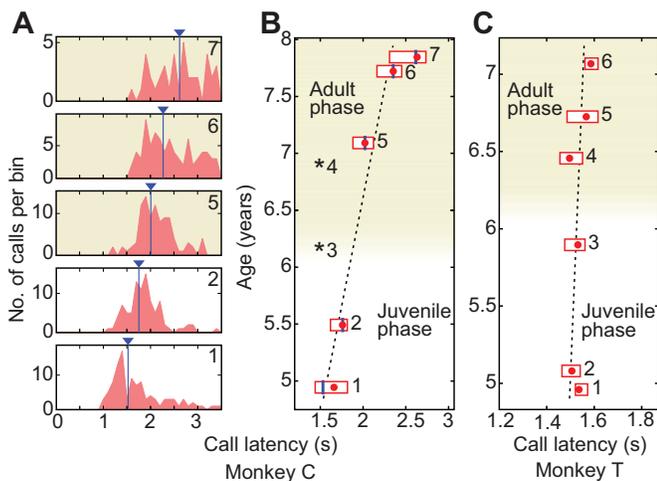


Fig. 3. Changes in call latency. (A) Examples of the distribution of median call latency in single sessions for monkey C during go trials in epochs 1, 2, 5, 6 and 7 (bin width 100 ms; call latency was not measured during epochs 3 and 4; see Materials and methods for details). Blue vertical lines indicate the median latency of the sessions. (B,C) Distribution of median call latency in the different epochs for monkey C (B) and monkey T (C). Colored dots inside boxes indicate medians; lower and upper margins of boxes represent the first and third quartile, respectively. Blue vertical bars in B are medians of the examples shown in A. Latencies of epoch 8 of monkey C (B) are not depicted because of low call numbers within the sessions; latencies of epoch 7 of monkey T (C) are not presented because of the absence of vocal behavior. *Epochs in which call latencies could not be determined in B (see Materials and methods for details). The shaded background indicates the adult period (Fleagle, 2013).

with microelectrodes in the prefrontal cortex (PFC) of both monkeys (Hage and Nieder, 2013, 2015), but we exclude the possibility that recordings caused damage that would have left the monkeys unable to vocalize on command. We have never witnessed a decline of any cognitive function as a result of PFC recordings, and post-mortem histological examination of other monkey brains has never showed damage to the tissue resulting from recordings. Furthermore, both monkeys have successfully been re-trained on other demanding tasks, and there was no indication whatsoever that the monkeys had suffered from disturbance of cognitive control functions. In fact, we argue that the visual DMS task that both monkeys successfully performed after they ceased to vocalize volitionally is more demanding than the cued vocalization (CV) task. In contrast to the CV task, the DMS task required discrimination of both sample and test stimuli (not just simple detection of a go stimulus) and memorization of a sample image over a delay period (which was entirely missing in the CV task). This is another indication that the monkeys were fully intact. Finally, we think it is highly unlikely that a putative worsened coordination between the manual and oral domains over development (the monkeys needed to grab a bar while vocalizing) might have caused the observed effects, given that hand movements and vocalizations were temporally disparate. Because the observed decline in volitional call behavior correlated with the transition of the monkeys from juvenile phases to adulthood, our findings can therefore best be reconciled with a maturation process. We suspect that early in ontogeny, the monkeys' neural central executive was still connected with the vocal motor network, thus allowing rudimentary cognitive control over call behavior. This cognitive control of vocal behavior was lost when the monkeys reached adulthood, pointing to developmental reorganization in the brain of these monkeys.

Using the identical task protocol, we previously reported a neuronal correlate of the monkeys' ability to initiate calls in response to the detection of an arbitrary visual stimulus (Hage and Nieder, 2013). Single neurons in the monkey homolog of Broca's area (Brodmann area 44 and 45) in the lateral PFC specifically signaled the preparation of instructed vocalizations, but not of spontaneous calls (Hage and Nieder, 2013). We hypothesize that these neurons of the PFC (which is generally associated with the brain's cognitive control center) connect the brain's executive with the vocal motor network early in primate ontogeny (Ackermann et al., 2014) as an obligate network for executive control on vocal output (Miller and Cohen, 2001). The anatomical substrate of this juvenile capability might be found in the excessive synaptic connections and dendritic spines particularly found in the PFC of human and non-human primates that are initially overproduced to about two times the adult number before being pruned during puberty to reach the adult level at the onset of adolescence (Petanjek et al., 2011; Bourgeois et al., 1994; Huttenlocher and Dabholkar, 1997; Dehaene and Cohen, 2007). This neoteny of brain structures in the PFC could be mediated by genes related to the development of the prefrontal cortex that show a correspondingly excessive, neotenic expression in humans relative to chimpanzees and rhesus macaques (Somel et al., 2013). Our hypothesis predicts that neural connections between the executive functioning networks in PFC and the brain's vocal motor network, which exist in juvenile monkeys, are decoupled during adolescence and are lost in adult monkeys. If true, such a finding would strengthen the neoteny hypothesis of human evolution (Gould, 1977) and explain aspects of human language evolution.

It is widely acknowledged that adolescence is associated with considerable reorganization of the brain. But what could cause the

loss of volitional vocalizations? Activity-dependent pruning of connections via elimination of excessive synapses is thought to play a major role in sculpting circuits and connections during ontogeny. However, because the brain networks to produce vocalizations were in use and of considerable behavioral relevance for our monkeys, the loss of this function would be difficult to reconcile with activity-dependent elimination of synapses. However, even without activity-dependent plasticity, the brain undergoes considerable reorganization during adolescence that serves a variety of other, possibly competing functions. For instance, hormonal changes associated with sexual maturation contribute to adolescent-typical behavioral changes that necessarily have an impact on large-scale networks. Functions beneficial during childhood may become inhibited during adulthood. In addition, changes of the highly interconnected brain in one area may in turn constrain the maintenance of other functions. Moreover, synaptic elimination during adolescence probably involves adjustment of the excitatory/inhibitory balance on individual neurons and within networks, given that excitatory synapses are selectively degenerated whereas inhibitory synapses are spared (Rakic et al., 1986). We speculate that the causes of the loss of brain circuits and networks for voluntary vocalizations are related to one (or several) of the non-activity-related elimination processes occurring in the maturing brain.

Our study emphasizes one of the rare cases of commonality between the human language system and non-human primate communication systems, namely the (developmentally restricted) ability to cognitively control vocalizations. It suggests that one important aspect of flexible communication is grounded in the primate lineage and could be exploited during the emergence of functional flexibility of prelinguistic vocalizations of human infants (Oller et al., 2013). As a phylogenetic pre-adaptation, volitional control of vocal utterances would be a crucial subcomponent in the complex multi-component system 'human language' and instrumental for all higher level linguistic characteristics emerging in human development, such as semantic compositionality or the grasp and mastering of a symbol system (Deacon, 1997; Nieder, 2009). Our behavioral study suggests an expansion of the juvenile period during ontogeny as one of the key evolutionary events in the evolution of language.

Acknowledgements

We thank two anonymous reviewers for helpful comments on a previous version of the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

S.R.H. and A.N. designed the study, interpreted the data and wrote the manuscript. S.R.H. and N.G. performed experiments and analyzed the data.

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 frame work of the Excellence Initiative EXC 307).

References

- Ackermann, H., Hage, S. R. and Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behav. Brain Sci.* **37**, 529–546.
- Altmann, J. (1974). Observational study of behavior: sampling methods. *Behaviour* **49**, 227–266.
- Arnold, K. and Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature* **441**, 303.
- Balter, M. (2010). Animal communication helps reveal roots of language. *Science* **328**, 969–971.

- Bourgeois, J.-P., Goldman-Rakic, P. S. and Rakic, P.** (1994). Synaptogenesis in the prefrontal cortex of rhesus monkeys. *Cereb. Cortex* **4**, 78-96.
- Carroll, S. B.** (2003). Genetics and the making of *Homo sapiens*. *Nature* **422**, 849-857.
- Coudé, G., Ferrari, P. F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., Monti, F., Rozzi, S. and Fogassi, L.** (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS ONE* **6**, e26822.
- Deacon, T. W.** (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: W. W. Norton.
- Deacon, T. W.** (2010). A role for relaxed selection in the evolution of the language capacity. *Proc. Natl. Acad. Sci. USA* **107**, 9000-9006.
- Dehaene, S. and Cohen, L.** (2007). Cultural recycling of cortical maps. *Neuron* **56**, 384-398.
- Fleagle, J. G.** (2013). *Primate Adaptation and Evolution*. Waltham, MA: Academic Press.
- Ghazanfar, A. A.** (2008). Language evolution: neural differences that make a difference. *Nat. Neurosci.* **11**, 382-384.
- Gould, S. J.** (1977). *Ontogeny and Phylogeny*. Cambridge, MA: Harvard University Press.
- Green, D. M. and Swets, J. A.** (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Hage, S. R. and Nieder, A.** (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* **4**, 2409.
- Hage, S. R. and Nieder, A.** (2015). Audio-vocal interaction in single neurons of the monkey ventrolateral prefrontal cortex. *J. Neurosci.* **35**, 7030-7040.
- Hage, S. R., Gavrilov, N. and Nieder, A.** (2013). Cognitive control of distinct vocalizations in rhesus monkeys. *J. Cogn. Neurosci.* **25**, 1692-1701.
- Hage, S. R., Ott, T., Eiselt, A.-K., Jacob, S. N. and Nieder, A.** (2014). Ethograms indicate stable well-being during prolonged training phases in rhesus monkeys used in neurophysiological research. *Lab. Anim.* **48**, 82-87.
- Hammerschmidt, K. and Fischer, J.** (2008). Constraints in primate vocal production. In *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication* (ed. D. K. Oller and U. Griebel), pp. 93-121. Cambridge, MA: MIT Press.
- Hammerschmidt, K., Jürgens, U. and Freudenstein, T.** (2001). Vocal development in squirrel monkeys. *Behaviour* **138**, 1179-1204.
- Huttenlocher, P. R. and Dabholkar, A. S.** (1997). Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167-178.
- Jürgens, U.** (2002). Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* **26**, 235-258.
- Koda, H., Oyakawa, C., Kato, A. and Masataka, N.** (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour* **144**, 681-692.
- Martin, P. and Bateson, P.** (1993). *Measuring Behaviour: An Introductory Guide*. Cambridge: University Press.
- Miller, E. K. and Cohen, J. D.** (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167-202.
- Nieder, A.** (2009). Prefrontal cortex and the evolution of symbolic reference. *Curr. Opin. Neurobiol.* **19**, 99-108.
- Oller, D. K.** (2000). *The Emergence of the Speech Capacity*. New York: Psychology Press.
- Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L. and Bakeman, R.** (2013). Functional flexibility of infant vocalization and the emergence of language. *Proc. Natl. Acad. Sci. USA* **110**, 6318-6323.
- Petanjek, Z., Judaš, M., Šimić, G., Rašin, M. R., Uylings, H. B. M., Rakic, P. and Kostovic, I.** (2011). Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc. Natl. Acad. Sci. USA* **108**, 13281-13286.
- Pistorio, A. L., Vintch, B. and Wang, X.** (2006). Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *J. Acoust. Soc. Am.* **120**, 1655.
- Rakic, P., Bourgeois, J.-P., Eckenhoff, M. F., Zecevic, N. and Goldman-Rakic, P. S.** (1986). Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science* **232**, 232-235.
- Seyfarth, R. M. and Cheney, D. L.** (1986). Vocal development in vervet monkeys. *Anim. Behav.* **34**, 1640-1658.
- Seyfarth, R. M. and Cheney, D. L.** (2010). Primate vocal communication. In *Primate Neuroethology* (ed. M. Platt and A. A. Ghazanfar), pp. 84-97. New York: Oxford Univ. Press.
- Somel, M., Liu, X. and Khaitovich, P.** (2013). Human brain evolution: transcripts, metabolites and their regulators. *Nat. Rev. Neurosci.* **14**, 112-127.
- Sutton, D., Larson, C., Taylor, E. M. and Lindeman, R. C.** (1973). Vocalization in rhesus monkeys: Conditionability. *Brain Res.* **52**, 225-231.
- Sutton, D., Larson, C. and Lindeman, R. C.** (1974). Neocortical and limbic lesion effects on primate phonation. *Brain Res.* **71**, 61-75.
- Sutton, D., Trachy, R. E. and Lindeman, R. C.** (1985). Discriminative phonation in macaques: effects of anterior mesial cortex damage. *Exp. Brain Res.* **59**, 410-413.
- Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P. and Ghazanfar, A. A.** (2015). The developmental dynamics of marmoset monkey vocal production. *Science* **349**, 734-738.
- Trachy, R. E., Sutton, D. and Lindeman, R. C.** (1981). Primate phonation: anterior cingulate lesion effects on response rate and acoustical structure. *Am. J. Primatol.* **1**, 43-55.