



Published in final edited form as:

Brain Behav Evol. 2013 ; 81(4): 219–225. doi:10.1159/000351203.

Rhesus macaques recognize unique multi-modal face-voice relations of familiar individuals and not of unfamiliar ones

Holly M. Habbershon¹, Sarah Z. Ahmed², and Yale E. Cohen²

¹University Laboratory Animal Resources Perelman School of Medicine at the University of Pennsylvania Philadelphia, PA 19104

²Department of Otorhinolaryngology Perelman School of Medicine at the University of Pennsylvania Philadelphia, PA 19104

Abstract

Communication signals in non-human primates are inherently multi-modal. However, for laboratory-housed monkeys, there is relatively little evidence in support of the use of multi-modal communication signals in individual recognition. Here, we used a preferential-looking paradigm to test whether laboratory-housed rhesus could “spontaneously” (i.e., in the absence of operant training) use multi-modal communication stimuli to discriminate between known conspecifics. The multi-modal stimulus was a silent movie of two monkeys vocalizing and an audio file of the vocalization from one of the monkeys in the movie. We found that the gaze patterns of those monkeys that knew the individuals in the movie were reliably biased toward the individual that did not produce the vocalization. In contrast, there was not a systematic gaze pattern for those monkeys that did not know the individuals in the movie. These data are consistent with the hypothesis that laboratory-housed rhesus can recognize and distinguish between conspecifics based on auditory and visual communication signals.

Keywords

rhesus; communication; vocalization; multi-modal; auditory; facial cue

Introduction

Non-human primates live in a complex social environment [Cheney et al. 1986; Partan and Marler 1999]. Consequently, it is likely important for them to be able to discriminate between familiar and unfamiliar conspecifics, recognize kin, and determine gender and social status [Cheney and Seyfarth 1990; Cheney and Seyfarth 1999; Cheney et al. 1986; Fugate et al. 2008; Gothard et al. 2009]. These abilities may be used by non-human primates to help them establish and maintain dominance relationships, non-kin alliances, and kin-biased behaviors [Cheney et al. 1986].

Correspondence: Dr. Yale Cohen Department of Otorhinolaryngology: Head and Neck Surgery Perelman School of Medicine at the University of Pennsylvania 3400 Spruce - 5 Ravdin Philadelphia, PA USA ycohen@mail.med.upenn.edu Phone: 1 215 898 7504 Fax: 1 215 898 9994.

A rich literature has demonstrated that these social-cognitive abilities originate from the ability of non-human primates to extract information from their sensory environment, such as the information that is transmitted by communication signals. For example, individuation is based, in part, on the ability of primates to discriminate between the unique timbres of different monkey's vocalizations [Cheney and Seyfarth 1980, 1999; Fischer 2004; Parr 2004; Rendall et al. 2000; Rendall et al. 1999; Snowdon and Cleveland 1980]. Facial cues and facial expressions also provide important information that allow monkeys to cognitively maneuver through their social environment [Deaner et al. 2005; Ghazanfar and Logothetis 2003; Gothard et al. 2009; Gothard et al. 2004; Parr 2004; Parr et al. 2000; Pascalis and Bachevalier 1998].

Although communication signals in non-human primates are inherently multi-modal [Ghazanfar and Schroeder 2006; Partan and Marler 1999], we know relatively little about how these multi-modal signals are used to discriminate between conspecifics. The studies that have addressed this issue have found that non-human primates can use multi-modal signals to discriminate between conspecifics (i.e., *A is different from B*) [Adachi and Hampton 2011; Evans et al. 2005; Ghazanfar and Logothetis 2003; Ghazanfar et al. 2007; Parr 2004]. There is also evidence to support the hypothesis that non-human primates can use multi-modal information to recognize different individuals (i.e., *this is A and this is B*) [Sliwa et al. 2011].

However, several issues still remain. For example, can previous field findings be corroborated under more controlled laboratory conditions? Moreover, since the laboratory setting is an impoverished environment (e.g., lack of extensive social interactions) relative to a rhesus' normal habitat, it is not clear whether monkeys recognize their conspecifics in a manner comparable to their free-roaming counterparts. Simultaneously, we wanted to test whether the results from a previous laboratory study [Sliwa et al. 2011] that utilized operant training to test for multi-modal recognition could be corroborated in the absence of operant training (i.e., "spontaneously") [Gifford et al. 2003].

Here, we addressed this question by having laboratory-housed rhesus participate in a preferential-looking paradigm [Fantz 1958, 1963]. In this paradigm, we presented a silent movie of two monkeys vocalizing and tested whether a monkey's gaze was modulated by the concomitant presentation of a vocalization of one of the monkeys in the movie. We found that the gaze patterns of those monkeys that knew the individuals in the movie were reliably biased toward the individual that did not produce the vocalization in the audio file. In contrast, there was not a systematic gaze pattern for those monkeys that did not know the individuals in the movie. These data are consistent with the hypothesis that laboratory-housed rhesus can recognize and distinguish between conspecifics based on multi-modal (auditory and visual) communication signals.

Methods

Subjects

Fourteen laboratory-housed adult male rhesus macaques (*Macaca mulatta*, 3 - 13 years old) were tested using a preferential-looking paradigm [Fantz 1958, 1963]. The Group-1 subjects

($n = 7$) were housed in one vivarium along with 5 other monkeys. Group-2 subjects ($n = 7$) were housed in a different vivarium along with 4 other monkeys. Each of the Group-1 and Group-2 monkeys was housed in their respective vivarium for at least 2.5 years with comparable housing and enrichment conditions. Of particular relevance to this study, each monkey had comparable multi-modal (visual and auditory) contact with all of the other monkeys in his respective vivarium. For the Group-1 subjects, this contact included the two monkeys (A and B) from which video and audio recordings were collected; see **Stimuli**. Critically, the monkeys in Group 1 and Group 2 were not in visual or auditory contact with one another. Thus, the Group-1 monkeys were the familiar group, whereas the Group-2 monkeys were the unfamiliar group. All experimental procedures were in accordance with the University of Pennsylvania's Institutional Animal Care and Use Committee.

Stimuli

Video and audio recordings of two monkeys (A and B) from Group-1 producing *grunt* vocalizations were recorded with a high-definition camcorder (Canon HFS30HD) while each monkey sat in an enclosed perch; this perch allowed a monkey to freely turn his head and body. Monkey A and monkey B were adult males (ages: 8 and 9 years respectively) and sizes (each weighed approximately 18 kg) and were housed with the Group-1 monkeys; for clarity, we will generally refer to monkey A and monkey B as the “stimulus” monkeys in order to differentiate them from the subject monkeys (see **Subjects**). The camcorder was focused only on the stimulus monkey's face and head. Video-editing software (Adobe Premiere Elements 9) extracted 1-s video segments of each stimulus monkey *grunting*. Background audio noise was removed with Adobe Audition 5. From each stimulus monkey, we obtained multiple video and audio recordings of their vocalizations. An example image and a spectrogram from each stimulus monkey are shown in Figure 1.

A silent 1-s video that contained the video clips of both stimulus-monkey A and stimulus-monkey B was paired and synchronized with a single instance of either stimulus-monkey A or stimulus-monkey B vocalizing: the video was edited so that the onsets of each monkey's mouth movements were synchronous. The video was also edited to ensure that the size of both stimulus monkeys was the same and that the video of each stimulus monkey, when displayed, was centered in one of the two video monitors (see **Stimulus Presentation and Testing Procedure** below). At the end of each 1-s video, there was a 2-s black screen without any audio. This process was repeated twenty times with different instances of video and audio combinations to generate a 60-s movie. In total, one hundred unique movies were generated. Each 1-s video was counterbalanced with respect to each stimulus monkey's appearance in each of the two video monitors and with respect to the vocalizing monkey (stimulus-monkey A versus stimulus-monkey B). That is, in any given video, the number of times that stimulus-monkey A was on the left and on the right was the same. Moreover, the number of times that an audio file from stimulus-monkey A or stimulus-monkey B was presented in a video was the same.

Stimulus Presentation and Testing Procedure

Each experimental session took place in a sound-attenuated room. Within this room, there were two video monitors (Acer L2016W) that were located at each subject monkey's eye

level (~132 cm). The distance between each monitor's centers was 65 cm or, relative to each subject monkey's position in the room, 25° (see Figure 2). Additionally, a speaker (Alpine SPR69C), which was covered to obscure it from a subject monkey's view, was placed midway between and at the same horizontal midline height as the two monitors. The camcorder was also placed above the speaker and recorded each subject monkey's behavioral responses (see **Data Analysis**).

At the beginning of an experimental session, a subject monkey was seated in his enclosed perch. Next, he was brought into the sound-attenuated room and placed in front of the two video monitors. Due to each subject monkey's participation in other visual- or auditory-perceptual experiments, he was comfortable and familiar with the perch and the environment of the sound-attenuated room. The only light in the room was a small lamp that illuminated the subject monkey's face for the camcorder.

Next, the subject monkey had a five-minute acclimation period in the darkened room. Following this acclimation period, an experimental session was initiated when the subject monkey's head was oriented toward the speaker. Once oriented, 5 randomly selected movies (see **Stimuli**) were presented to each subject monkey. Following the conclusion of the last movie, the subject monkey was given pieces of fresh fruit and returned to his home cage. Each of the Group-1 and Group-2 subject monkeys participated in one experimental session. The video recordings obtained from each session were analyzed off-line.

Data Analysis

From the video footage, we quantified “match time” and “non-match” time. Match time was the amount of time each subject monkey's eye position was directed toward the center of the monitor that showed the video of the subject monkey that produced the vocalization (e.g., video of stimulus-monkey A and vocalization of stimulus-monkey A). Non-match time was the amount of time each subject monkey's eye position was directed toward the center of the monitor that showed the video of the stimulus monkey that did not produce the vocalization (e.g., video of stimulus-monkey A and vocalization of stimulus-monkey B). The start of the match or non-match times began when a subject monkey's gaze was directed at the center of a video monitor and ended when his gaze oriented away from the video monitor or when he closed his eyes. Independent of the amount of head movement, the subject monkey's pupils had to be visible to be included as part of the match or non-match time. To validate the data analysis, in addition to two of the authors of this manuscript, two additional naïve observers, who were blind to the purpose of the videos and to the subject monkeys' identities, independently quantified each subject monkey's match and non-match times. The inter-observer correlation was 0.96.

Results

Each of the 14 subject monkeys participated in one experimental session, during which they viewed five randomly selected movies (see **Stimuli**). We found that the total amount of time (i.e., match time plus non-match time) that the subject-monkey Groups 1 and 2 watched each movie was not the same (two-factor ANOVA, movie presentation \times subject-monkey group [Group 1 and Group 2], $F = 4.87$, $p < 0.05$). Indeed, all of the subject monkeys

watched the first movie but watched subsequent movies to varying degrees. Indeed, some monkeys only watched the first movie. Further analyses indicated that there was not a reliable ($t = 1.8, p > 0.05$) difference between the total amount of time that the Group-1 subject monkeys (mean = 9.4 s, s.d. = 2.3 s) and Group-2 subject monkeys watched the first movie (mean = 6.6 s; s.d. = 3.2 s). This latter result is important because it indicates that neither (1) the familiarity of viewing known conspecifics for the Group-1 subject monkeys nor (2) the novelty of viewing unfamiliar conspecifics for the Group-2 subject monkeys biased the amount of time that the subject monkeys spent viewing the videos. Since the amount of time that the Group-1 and Group-2 subject monkeys watched the first movie was not reliably different and since some of the monkeys did not watch any movie but the first one, our subsequent analysis focused wholly on the subject monkeys' responses during the first movie.

The results of this analysis are shown in Figure 3. In this figure, the match time (i.e., the amount of the time a subject monkey oriented his gaze toward the video of the stimulus monkey that produced the vocalization) and non-match time (i.e., the amount of time that a subject monkey oriented his gaze toward the video of the stimulus monkey that did not produce the vocalization) are shown for each subject monkey in each group. As can be seen, most ($n = 6/7$) of the Group-1 subject monkeys (i.e., those who were familiar with the individuals in the videos) had the same behavior: that is, they had longer non-match times than match times. On average, the Group-1 subject monkeys had average non-match times of 7.1 s (s.d. = 2.2 s) and match times of 2.2 s (s.d. = 1.3). Both the average non-match time ($t = -3.7, df = 6, p < 0.05$) and the average match time ($t = 2.6, df = 6, p < 0.05$) were reliably different from chance (i.e., the average of the Group-1 subject monkeys match and non-match times). Furthermore, the Group-1 subject monkeys had, on average, reliably longer non-match times than match times ($t = -4.77, df = 6, p < 0.05$).

In contrast, a different behavior was observed for the Group-2 subject monkeys (i.e., those who were unfamiliar with the individuals in the videos). First, only 4 out of the 7 subject monkeys had longer non-match times than matches. Second, the average non-match times (4.1 s; s.d. = 1.9) and the average match times (2.6 s; s.d. = 2.3) of the Group-2 subject monkeys were not reliably different from chance (non-match time: $t = -0.68, df = 6, p > 0.05$; match time: $t = 0.79, df = 6, p > 0.05$); analogous to the Group-1 subject monkeys, chance was defined the average of the Group-2 subject monkeys match and non-match times). Finally, for the Group-2 subject monkeys, we could not identify a reliable difference between the non-match and match times ($t = -1.4, df = 6, p > 0.05$).

A direct comparison between the Group-1 and Group-2 subject monkeys yielded the following pattern of results. First, the Group-1 subject monkeys had reliably longer non-match times than the Group-2 subject monkeys ($t = 2.9, df = 6, p < 0.05$). However, we could not identify a reliable difference between the Group-1 and Group-2 subject monkeys' match times ($t = -0.34, df = 6, p > 0.05$).

Discussion

The results of this study suggest that laboratory-housed rhesus can differentiate between individuals based on auditory and visual information, adding to the rich literature on multi-modal representation of individuation in human and non-human animals [Adachi and Fujita 2007; Adachi and Hampton 2011; Campanella and Belin 2007; Kondo et al. 2012; Proops et al. 2009; Sliwa et al. 2011]. Two factors in our experimental design potentially limit interpretation of our findings. First, whereas both stimulus monkeys were the same age, sex, and weight, our findings may be attributed to differences in their dominance status. Our observation of the colony suggests that the two stimulus monkeys have equal status; however, we cannot conclusively determine their relative status. Second, since we only had two stimulus monkeys, generalizing our finding to individual recognition may be limited. Nevertheless, given these limitations, below, we interpret our data within the context of previous findings and discuss potential brain regions involved in multi-modal differentiation and recognition of con-specifics.

The findings of this study expand our understanding of primate cognition in two ways. First, we compared gaze patterns directed to multi-modal representations of known individuals with those toward unknown individuals. In human studies [Kamachi et al. 2003; Lander et al. 2007], it was shown that subjects can match unknown facial identity with unknown vocal identity. This ability, however, was not present when our subject monkeys were asked to do this with same (approximately) aged and sexed monkeys. The basis for this evolutionary difference is not known but may relate differences in how auditory and visual information are integrated by human and non-human primates [Kamachi et al. 2003; Lander et al. 2007]. Second, whereas visual attention directed toward stimuli representing familiar versus unfamiliar con-specifics have been studied using uni-modal stimuli [Mahajan et al. 2011; Pokorny and de Waal 2009; Schell et al. 2011], our study is the first, to our knowledge, to address this question with multi-modal stimuli. More specifically, we tested whether rhesus can recognize unique multi-modal facial-vocalization relationships rather than whether monkeys can sequentially cross-modally match a face to a vocalization (e.g., Sliwa et al., 2011 or Adachi and Hampton, 2011). This finding suggests that rhesus have, at least, an undividable representation of individuals. As a point of comparison, in human infants, the ability to create such an undividable representation of an individual precedes their ability to sequentially match a face to a voice [Bahrack et al. 2005].

How then do we interpret our finding that the Group-1 subject monkeys had longer non-match times? We interpret our findings within the classic “violation of expectation” findings that were first demonstrated in human infants by Spelke and colleagues [Spelke 1990, 1994; Spelke et al. 1992]. Specifically, in our study, when one of the subject monkeys heard a vocalization, we hypothesized that the monkey formed a representation [Adachi and Fujita 2007; Adachi and Hampton 2011; Adachi et al. 2007; Carey 2010; Johnston and Bullock 2001; Sliwa et al. 2011] of the individual eliciting the call. This representation created an expectation of what the listener should see when he oriented his gaze toward the caller. However, when the subject monkey did not see the individual that he expected to elicit the vocalization, he was surprised (a violation of the expectation), therefore he looked longer (i.e., had longer non-match times).

Since vocalizations are used in different contexts, a reasonable hypothesis may be that a monkey's ability to use multi-modal signals is restricted to certain types of vocalizations. For example, *coos* are harmonically structured sounds that are markedly more distinctive by vocalizer than either *grunts* or *noisy screams*, and that spectral-patterning measures related to vocal tract filtering effects are the most reliable markers of individual identity [Rendall et al. 1998]. In contrast, *grunts*, which are pulsatile noisy vocalizations and are classified at a lower but above-chance rate. Consequently, rhesus commonly see and hear con-specifics *grunting* when they are in close contact but see and hear con-specifics *cooing* at others at a distance. However, this hypothesis does not seem to be correct: monkeys can also use the multi-modal information provided in vocalizations such as *coos* [Adachi and Hampton 2011; Ghazanfar and Logothetis 2003; Sliwa et al. 2011], which are produced when monkeys are not in direct proximity [Ghazanfar et al. 2001; Ghazanfar et al. 2007; Hauser and Marler 1993]. Thus, the use of multi-modal signals to differentiate and recognize conspecifics is not limited to contextually specific vocalizations. Instead, it is a process that depends on the multi-modal signal features that allow for individuation (e.g., vocalization timbre) [Cheney and Seyfarth 1980, 1999; Fischer 2004; Parr 2004; Rendall et al. 2000; Rendall et al. 1999; Snowden and Cleveland 1980].

Finally, what brain areas contribute to multi-modal recognition? Likely brain areas that contribute to multi-modal recognition include the core and lateral belt regions of auditory cortex, as well as the ventrolateral prefrontal cortex. These brain regions are part of the ventral auditory pathway; this pathway is specialized for processing information about stimulus identity [Miller and Cohen 2010; Rauschecker and Tian 2000; Romanski and Averbeck 2009]. The core and lateral belt regions of the auditory cortex have been shown to integrate facial and vocal signals through the enhancement and suppression of field potentials [Ghazanfar et al. 2005] and likely play a fundamental role in the multi-modal integration of communication signals. The prefrontal cortex also appears to play a key role in the integration of multi-modal communication signals [Romanski and Hwang 2012] that can guide adaptive behavior [Miller and Cohen 2001]. Indeed, neural activity in both the auditory cortex and the prefrontal cortex are differentially modulated when facial and vocal signals match versus when they do not match [Ghazanfar et al. 2005; Romanski and Hwang 2012]. Moreover, inactivation of the prefrontal cortex impairs the ability of rhesus to correctly match a vocalization with its corresponding facial image [Plakke et al. 2012]. It is, therefore, likely that our behavioral findings rely on the neural computations that occur in one or more of the brain regions in the ventral auditory pathway.

Acknowledgments

We would like to thank Robert Seyfarth, Asif Ghazanfar, and Heather Hersh for helpful comments on the preparation of this manuscript. We also want to thank Harry Shirley and Jean Zweigle for superb animal care. This research was supported by grants from the NIDCD to YEC.

Reference

- Adachi I, Fujita K. Cross-modal representation of human caretakers in squirrel monkeys. *Behavioural processes*. 2007; 74:27–32. [PubMed: 17046173]
- Adachi I, Kuwahata H, Fujita K. Dogs recall their owner's face upon hearing the owner's voice. *Animal Cognition*. 2007; 10:17–21. [PubMed: 16802145]

- Adachi I, Hampton RR. Rhesus monkeys see who they hear: Spontaneous cross-modal memory for familiar conspecifics. *PLoS One*. 2011; 6:e23345. [PubMed: 21887244]
- Bahrack LE, Hernandez-Reif M, Flom R. The development of infant learning about specific face-voice relations. *Dev Psychobiol*. 2005; 41:541–552.
- Campanella S, Belin P. Integrating face and voice in person perception. *Trends in cognitive sciences*. 2007; 11:535–543. [PubMed: 17997124]
- Carey, S. *The origin of concepts*. Oxford University Press; New York: 2010.
- Cheney DL, Seyfarth RM. Vocal recognition in free ranging vervet monkeys. *Anim Behav*. 1980; 28:362–367.
- Cheney DL, Seyfarth RM, Smuts B. Social relationships and social cognition in nonhuman primates. *Science*. 1986; 234:1361–1366. [PubMed: 3538419]
- Cheney, DL.; Seyfarth, RM. *How monkeys see the world*. University of Chicago Press; Chicago, IL: 1990.
- Cheney DL, Seyfarth RM. Recognition of other individuals' social relationships by female baboons. *Anim Behav*. 1999; 58:67–75. [PubMed: 10413542]
- Deaner RO, Khera AV, Platt ML. Monkeys pay per view: Adaptive valuation of social images by rhesus macaques. *Current Biology*. 2005; 15:543–548. [PubMed: 15797023]
- Evans TA, Howell S, Westergaard GD. Auditory-visual cross-modal perception of communicative stimuli in tufted capuchin monkeys (*cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*. 2005:399–406. [PubMed: 16248726]
- Fantz RL. Pattern vision in young infants. *Psych Rec*. 1958; 8:43–47.
- Fantz RL. Pattern vision in newborn infants. *Science*. 1963; 140:296–297. [PubMed: 17788054]
- Fischer J. Emergence of individual recognition in young macaques. *Anim Behav*. 2004; 67:655–661.
- Fugate JM, Gouzoules H, Nygaard LC. Recognition of rhesus macaque (*macaca mulatta*) noisy screams: Evidence from conspecifics and human listeners. *Am J Primatol*. 2008; 70:594–604. [PubMed: 18318042]
- Ghazanfar AA, Smith-Rohrberg D, Hauser MD. The role of temporal cues in rhesus monkey vocal recognition: Orienting asymmetries to reversed calls. *Brain Behav Evol*. 2001; 58:163–172. [PubMed: 11910173]
- Ghazanfar AA, Logothetis NK. Neuroperception: Facial expressions linked to monkey calls. *Nature*. 2003; 424:937–938. [PubMed: 12827188]
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*. 2005; 25:5004–5012. [PubMed: 15901781]
- Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? *Trends Cog Sci*. 2006; 10:278–285.
- Ghazanfar AA, Turesson HK, Maier JX, van Dinther R, Patterson RD, Logothetis NK. Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology*. 2007; 17:425–430. [PubMed: 17320389]
- Gifford GW 3rd, Hauser MD, Cohen YE. Discrimination of functionally referential calls by laboratory-housed rhesus macaques: Implications for neuroethological studies. *Brain Behav Evol*. 2003; 61:213–224. [PubMed: 12784058]
- Gothard KM, Erickson CA, Amaral DG. How do rhesus monkeys (*macaca mulatta*) scan faces in a visual paired comparison task? *Animal Cognition*. 2004; 7:25–36. [PubMed: 14745584]
- Gothard KM, Brooks KN, Peterson MA. Multiple perceptual strategies used by macaque monkeys for face recognition. *Animal Cognition*. 2009; 12:155–167. [PubMed: 18787848]
- Hauser MD, Marler P. Food-associated calls in rhesus macaques (*macaca mulatta*) 1. Socioecological factors influencing call production. *Behav Ecol*. 1993; 4:194–205.
- Johnston RE, Bullock TA. Individual recognition by use of odors in golden hamsters: The nature of individual representation. *Anim Behav*. 2001:545–557.
- Kamachi M, Hill H, Lander K, Vatikiotis-Bateson E. “Putting the face to the voice”: Matching identity across modality. *Curr Biol*. 2003; 13:1709–1714. [PubMed: 14521837]

- Kondo N, Izawa E, Watanabe S. Crows cross-modally recognize group members but not non-group members. *Proceedings Biological sciences / The Royal Society*. 2012; 279:1937–1942. [PubMed: 22217722]
- Lander K, Hill H, Kamachi M, Vatikiotis-Bateson E. It's not what you say but the way you say it: Matching faces and voices. *J Exp Psychol Hum Percept Perform*. 2007; 33:905–914. [PubMed: 17683236]
- Mahajan N, Martinez MA, Gutierrez NL, Diesendruck G, Banaji MR, Santos LR. The evolution of intergroup bias: Perceptions and attitudes in rhesus macaques. *Journal of personality and social psychology*. 2011; 100:387–405. [PubMed: 21280966]
- Miller, CT.; Cohen, YE. Vocalization processing. In: Ghazanfar, A.; Platt, ML., editors. *Primate neuroethology*. Oxford University Press; Oxford, UK: 2010. p. 237-255.
- Miller E, Cohen JD. An integrative theory of prefrontal cortex function. *Ann Rev Neurosci*. 2001; 24:167–202. [PubMed: 11283309]
- Parr LA, Winslow JT, Hopkins WD, de Waal FB. Recognizing facial cues: Individual discrimination by chimpanzees (*pan troglodytes*) and rhesus monkeys (*macaca mulatta*). *J Comp Psychol*. 2000; 114:47–60. [PubMed: 10739311]
- Parr LA. Perceptual biases for multimodal cues in chimpanzee (*pan troglodytes*) affect recognition. *Animal Cognition*. 2004; 7:171–178. [PubMed: 14997361]
- Partan S, Marler P. Communication goes multimodal. *Science*. 1999; 283:1272–1273. [PubMed: 10084931]
- Pascalis O, Bachevalier J. Face recognition in primates: A cross-species study. *Behavioral Processes*. 1998; 43:87–96.
- Plakke, B.; Hwang, J.; Diltz, MD.; Romanski, LM. Inactivation of ventral prefrontal cortex impairs audiovisual working memory. Program 87810 2012 Neuroscience Meeting Planner; New Orleans, LA: Society for Neuroscience; 2012.
- Pokorny JJ, de Waal FB. Monkeys recognize the faces of group mates in photographs. *Proc Natl Acad Sci USA*. 2009; 106:21539–21543. [PubMed: 19966230]
- Proops L, McComb K, Reby D. Cross-modal individual recognition in domestic horses (*equus caballus*). *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:947–951. [PubMed: 19075246]
- Rauschecker JP, Tian B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci USA*. 2000; 97:11800–11806. [PubMed: 11050212]
- Rendall D, Owren MJ, Rodman PS. The role of vocal tract filtering in identity cueing in rhesus monkey (*macaca mulatta*) vocalizations. *The Journal of the Acoustical Society of America*. 1998; 103:602–614. [PubMed: 9440345]
- Rendall D, Seyfarth RM, Cheney DL, Owren MJ. The meaning and function of grunt variants in baboons. *Anim Behav*. 1999; 57:583–592. [PubMed: 10196047]
- Rendall D, Cheney DL, Seyfarth RM. Proximate factors mediating “contact” calls in adult female baboons (*papio cynocephalus ursinus*) and their infants. *Journal of Comparative Psychology*. 2000; 114:36–46. [PubMed: 10739310]
- Romanski LM, Averbeck BB. The primate cortical auditory system and neural representation of conspecific vocalizations. *Ann Rev Neurosci*. 2009:315–346. [PubMed: 19400713]
- Romanski LM, Hwang J. Timing of individual inputs to the prefrontal cortex and multisensory integration. *Neuroscience*. 2012
- Schell A, Rieck K, Schell K, Hammerschmidt K, Fischer J. Adult but not juvenile barbary macaques spontaneously recognize group members from pictures. *Animal Cognition*. 2011; 14:503–509. [PubMed: 21318387]
- Sliwa J, Duhamel JR, Pascalis O, Wirth S. Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:1735–1740. [PubMed: 21220340]
- Snowdon CT, Cleveland J. Individual recognition of contact calls by pygmy marmosets. *Anim Behav*. 1980; 33:272–283.
- Spelke ES. Principles of object perception. *Cognitive Science*. 1990; 14:29–56.

Spelke ES, Breinlinger K, Macomber J, Jacobson K. Origins of knowledge. *Psychology Review*. 1992; 99:605–632.

Spelke ES. Initial knowledge: Six suggestions. *Cogn*. 1994; 50:431–445.

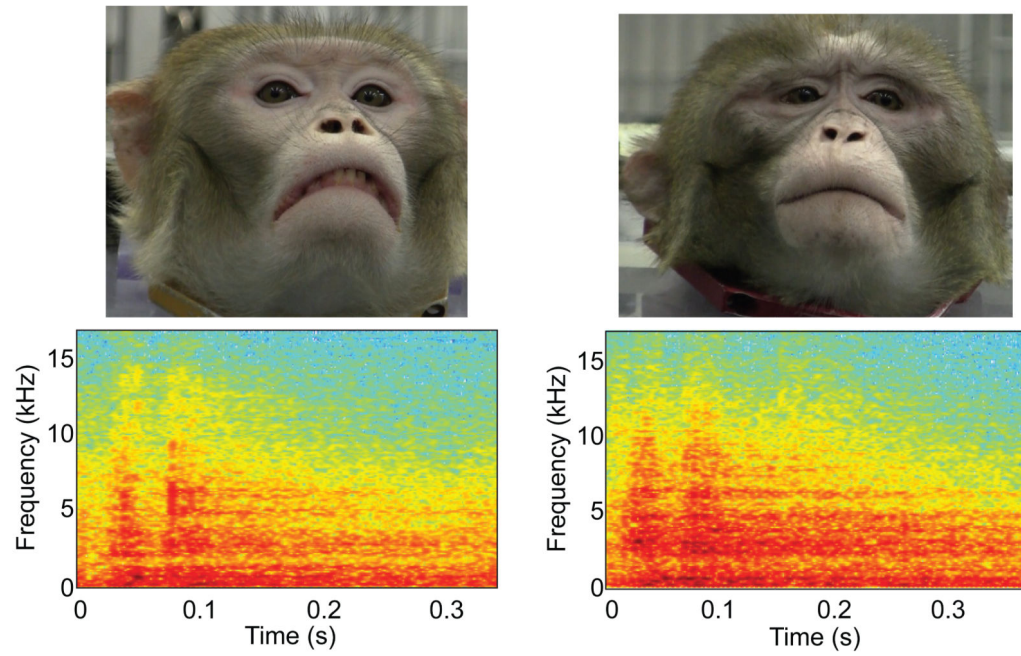


Figure 1. Visual and auditory stimuli. A snapshot of each stimulus monkey and a spectrogram from each monkey. Stimulus-monkey A is shown on the left, whereas stimulus-monkey B is shown on the right.

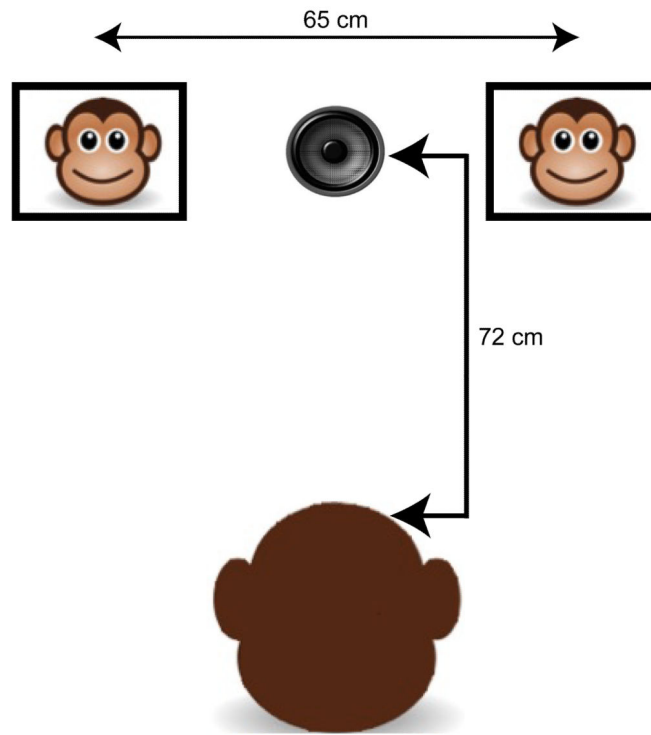


Figure 2.

Schematic of the experimental set-up in the sound-attenuated room showing placement of a monkey and the two video monitors. Two video monitors were located at each monkey's eye level (~132cm). The monitor-to-monitor separation was 65 cm or 25° , relative to the monkey's position in the room. A speaker, covered to obscure it from the monkey's view (shown uncovered in the figure), was placed midway and at the same height as the two monitors.

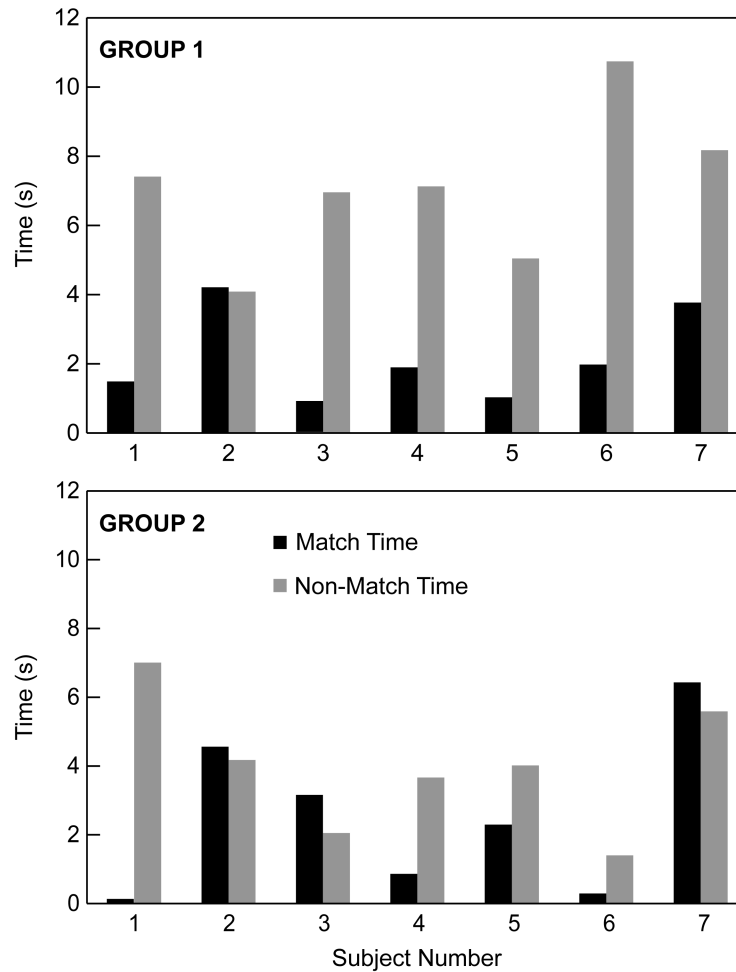


Figure 3. Match and non-match times. The match and non-match times for Group 1 (top) and Group 2 (bottom) subject monkeys are shown for each individual monkey. The black bars indicate match times and the grey bars indicate the non-match times for each subject monkey. 6 of the 7 Group-1 monkeys had longer non-match times than match times. 3 of the 7 Group-2 monkeys had longer non-match times than match times.