Chapter 9
The Default Mode of Primate Vocal Communication and Its Neural Correlates

Asif A. Ghazanfar

9.1 Introduction

It's been argued that the integration of the visual and auditory channels during human speech perception is the default mode of speech processing (Rosenblum, 2005). That is, speech perception is not a capacity that is 'piggybacked' on to auditory-only speech perception. Visual information from the mouth and other parts of the face is used by all perceivers and readily integrates with auditory speech. This integration is ubiquitous and automatic (McGurk and MacDonald, 1976) and is similar across all sighted individuals across all cultures (Rosenblum, 2008). The two modalities seem to be integrated even at the earliest stages of human cognitive development (Gogate et al., 2001; Patterson and Werker, 2003).

If multisensory speech is the default mode of perception, then this should be reflected both in the evolution of vocal communication and in the organization of neural processes related to communication. The purpose of this chapter is (1) to briefly describe the data that reveal that human speech is not uniquely multisensory, that in fact, the default mode of communication is multisensory in nonhuman primates as well and (2) to suggest that this mode of communication is reflected in the organization of the neocortex. By focusing on the properties of a presumptive unisensory region – the auditory cortex – I will argue that multisensory associations are not mediated solely through association areas, but are instead mediated through large-scale networks that include both 'lower' and 'higher' sensory areas.

9.2 Faces and Voices Are Inextricably Linked in Primates

Human and primate vocalizations are produced by coordinated movements of the lungs, larynx (vocal folds), and the supralaryngeal vocal tract (Fitch and Hauser,

A.A. Ghazanfar (✉)
Departments of Psychology and Ecology & Evolutionary Biology, Neuroscience Institute, Princeton University, Princeton, NJ 08540, USA
e-mail: asifg@princeton.edu

M.J. Naumer, J. Kaiser (eds.), *Multisensory Object Perception in the Primate Brain*,
DOI 10.1007/978-1-4419-5615-6_9,

1995; Ghazanfar and Rendall, 2008). The vocal tract consists of the column of air derived from the pharynx, mouth, and nasal cavity. In humans, speech-related vocal tract motion results in the predictable deformation of the face around the oral aperture and other parts of the face (Jiang et al., 2002; Yehia et al., 1998, 2002). Thus, facial motion is inextricably linked to the production of vocal sounds. For example, human adults automatically link high-pitched sounds to facial postures producing an /i/ sound and low-pitched sounds to faces producing an /a/ sound (Kuhl et al., 1991). In nonhuman primate vocal production, there is a similar link between acoustic output and facial dynamics. Different macaque monkey vocalizations are produced with unique lip configurations and mandibular positions and the motion of such articulators influences the acoustics of the signal (Hauser and Ybarra, 1994; Hauser et al., 1993). Coo calls, like /u/ in speech, are produced with the lips protruded, while screams, like the /i/ in speech, are produced with the lips retracted (Fig. 9.1). Thus, it is likely that many of the facial motion cues that humans use for speech-reading are present in other primates as well.

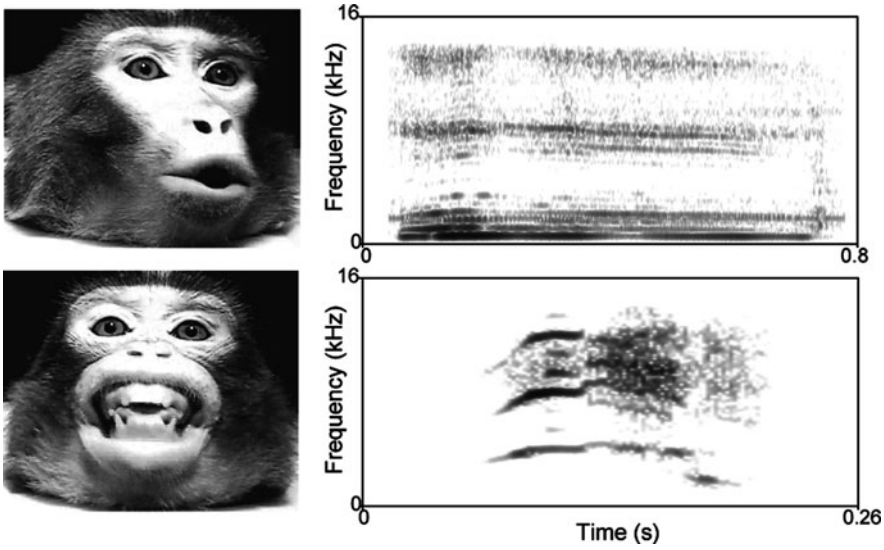

Fig. 9.1 Exemplars of the facial expressions produced concomitantly with vocalizations. Rhesus monkey coo and scream calls taken at the midpoint of the expressions with their corresponding spectrograms

The link between facial motion and vocalizations presents an obvious opportunity to exploit the concordance of both channels. Thus, it is not surprising that many primates other than humans recognize the correspondence between the visual and the auditory components of vocal signals. Rhesus and Japanese macaques (*Macaca mulatta and Macaca fuscata*), capuchins (*Cebus apella*), and chimpanzees (*Pan troglodytes*) (the only nonhuman primates tested thus far) all recognize

auditory–visual correspondences between their various vocalizations (Adachi et al., 2006; Evans et al., 2005; Ghazanfar and Logothetis, 2003; Izumi and Kojima, 2004; Parr, 2004). For example, rhesus monkeys readily match the facial expressions of 'coo' and 'threat' calls with their associated vocal components (Ghazanfar and Logothetis, 2003). Perhaps more pertinent, rhesus monkeys can also segregate competing voices in a chorus of coos, much as humans might with speech in a cocktail party scenario, and match them to the correct number of individuals seen cooing on a video screen (Jordan et al., 2005) (Fig. 9.2a). Finally, macaque monkeys use

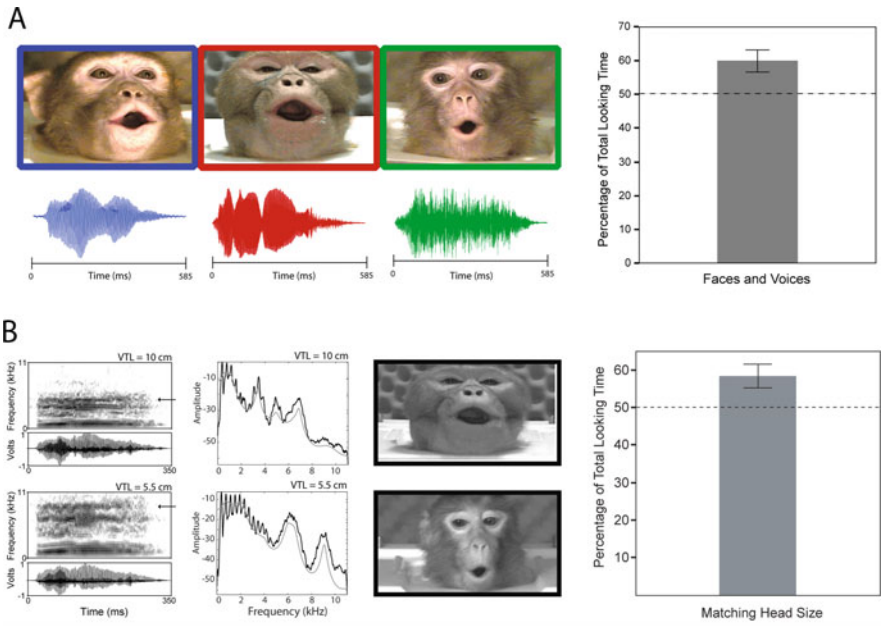

Fig. 9.2 Monkeys can match across modalities. To test this, we adopted the preferential-looking paradigm which does not require training or reward. In the paradigm, subjects are seated in front of two LCD monitors and shown two side-by-side digital videos, only one of which corresponds to the sound track heard through a centrally located speaker. A trial consists of two videos played in a continuous loop with one of the two sound tracks also played in a loop through the speaker. The dependent measure is percentage of total looking time to the match video. (**a**) Monkeys segregate coo vocalizations from different individuals and look to correct number of conspecific individuals displayed on the screen. Still frames extracted from a stimulus set along with their acoustic counterparts below. The bar graph shows the mean percentage (±SEM) of total looking time to the matching video display; chance is 50%. (**b**) A single coo vocalizations were synthesized to mimic large and small sounding individuals. Diagrams in the *left panels* show the spectrograms and waveforms of a coo vocalization re-synthesized with two different vocal tract lengths. The *arrow* in the spectrogram indicates the position of an individual formant which increases in frequency as the apparent vocal tract length decreases. In the *middle panels*, power spectra (*black line*) and linear predictive coding spectra (*gray lines*) for the long vocal tract length (10 cm, *top panel*) and short vocal tract length (5.5 cm, *bottom panel*). Still frames show the visual components of a large and small monkey. The bar graph shows the mean percentage of total time spent looking at the matching video display; the *dotted line* indicates chance expectation. Error bars are SEM

formants (i.e., vocal tract resonances) as acoustic cues to assess age-related body size differences among conspecifics (Ghazanfar et al., 2007) (Fig. 9.2b). They do so by linking across modalities the body size information embedded in the formant spacing of vocalizations (Fitch, 1997) with the visual size of animals who are likely to produce such vocalizations (Ghazanfar et al., 2007). Taken together these data suggest that humans are not at all unique in their ability to communicate across modalities. Indeed, as will be described below, vocal communication is a fully integrated *multi-sensorimotor* system with numerous similarities between humans and monkeys and in which auditory cortex may serve as a key node in a larger neocortical network.

9.3 Neocortical Bases for Integrating Faces and Voices

Although it is generally recognized that we and other animals use our different senses in an integrated fashion, we assume that, at the neural level, these senses are, for the most part, processed independently but then converge at critical nodes. This idea extends as far back as Leonardo da Vinci's (1452–1519) research into the neuroanatomy of the human brain. He suggested that there was an area above the pituitary fossa where the five senses converged (the 'sensu comune') (Pevsner, 2002). The basic tenet of neocortical organization has not changed to a large degree since da Vinci's time as it has long been argued that different regions of the cortex have different functions segregated according to sense modality. Some regions receive visual sensations, others auditory sensations, and still others tactile sensations (so and so forth, for olfaction and gustation). Each of these sensory regions is thought to send projections which converge on an 'association area' which then enables the association between the different senses and between the senses and the movement.

According to this line of thinking, the linking of vision with audition in the multisensory vocal perception described above would be attributed to the functions of association areas such as the superior temporal sulcus in the temporal lobe or the principal and intraparietal sulci located in the frontal and parietal lobes, respectively. Although these regions may certainly play important roles (see below), they are certainly not necessary for all types of multisensory behaviors (Ettlinger and Wilson, 1990), nor are they the sole regions for multisensory convergence (Driver and Noesselt, 2008; Ghazanfar and Schroeder, 2006). The auditory cortex, in particular, has many potential sources of visual inputs (Ghazanfar and Schroeder, 2006) and this is borne out in the increasing number of studies demonstrating visual modulation of auditory cortical activity (Bizley et al., 2007; Ghazanfar et al., 2005, 2008; Kayser et al., 2007, 2008; Schroeder and Foxe, 2002). Here I focus on those auditory cortical studies investigating face/voice integration specifically.

Neural activity in both the primary and the lateral belt regions of auditory cortex is influenced by the presence of a dynamic face (Ghazanfar et al., 2005, 2008). Monkey subjects viewing unimodal and bimodal versions of two

different species-typical vocalizations ('coos' and 'grunts') show both enhanced and suppressed local field potential (LFP) responses in the bimodal condition relative to the unimodal auditory condition (Ghazanfar et al., 2005). Consistent with evoked potential studies in humans (Besle et al., 2004; van Wassenhove et al., 2005), the combination of faces and voices led to integrative responses (significantly different from unimodal responses) in the vast majority of auditory cortical sites – both in the primary auditory cortex and in the lateral belt auditory cortex. The data demonstrated that LFP signals in the auditory cortex are capable of multisensory integration of facial and vocal signals in monkeys (Ghazanfar et al., 2005) and have subsequently been confirmed at the single-unit level in the lateral belt cortex as well (Ghazanfar et al., 2008) (Fig. 9.3).

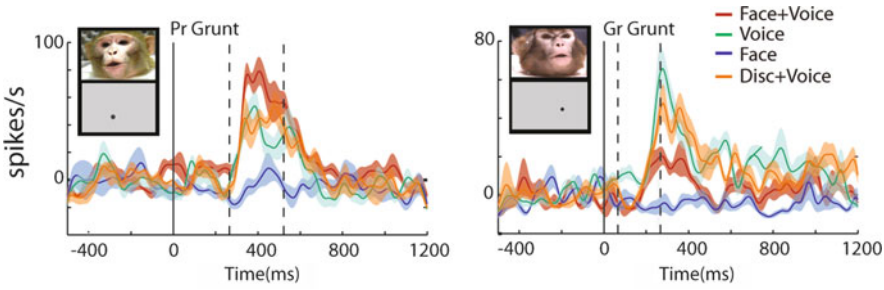

Fig. 9.3 Single neuron examples of multisensory integration of face+voice stimuli compared with disk+voice stimuli in the lateral belt area. The *left panel* shows an enhanced response when voices are coupled with faces, but no similar modulation when coupled with disks. The *right panel* shows similar effects for a suppressed response. *x*-axes show time aligned to onset of the face (*solid line*). *Dashed lines* indicate the onset and offset of the voice signal. *y*-axes depict the firing rate of the neuron in spikes per second. *Shaded regions* denote the SEM

To test the specificity of face/voice integrative responses, the dynamic faces were replaced with dynamic discs which mimicked the aperture and displacement of the mouth. In human psychophysical experiments, such artificial dynamic stimuli can still lead to enhanced speech detection, but not to the same degree as a real face (Bernstein et al., 2004; Schwartz et al., 2004). When cortical sites or single units were tested with dynamic discs, far less integration was seen when compared to the real monkey faces (Ghazanfar et al., 2005, 2008) (Fig. 9.3). This was true primarily for the lateral belt auditory cortex (LFPs and single units) and was observed to a lesser extent in the primary auditory cortex (LFPs only).

A comparison of grunt calls versus coo calls revealed that the former seemed to be over-represented relative to the latter. That is, grunt vocalizations elicited more enhanced multisensory LFP responses than did coo calls (Ghazanfar et al., 2005). As both coos and grunts are produced frequently in a variety of affiliative contexts and are broadband spectrally, the differential representation cannot be attributed to experience, valence, or the frequency tuning of neurons. One remaining possibility

is that this differential representation may reflect a behaviorally relevant distinction, as coos and grunts differ in their direction of expression and range. Coos are generally contact calls rarely directed toward any particular individual. In contrast, grunts are often directed toward individuals in one-on-one situations, often during social approaches as in baboons and vervet monkeys (Cheney and Seyfarth, 1982; Palombit et al., 1999). Given their production at close range and context, grunts may produce a stronger face/voice association than coo calls. This distinction appeared to be reflected in the pattern of significant multisensory responses in auditory cortex; that is, this multisensory bias toward grunt calls may be related to the fact that the grunts (relative to coos) are often produced during intimate, one-to-one social interactions. That said, much more work needs to be done to explore whether these multisensory differences are simply due to greater numbers of auditory neurons responding to grunts in general (something that the LFP signal cannot tell us) or whether there is truly something unique about the face/voice integration process for this vocalization.

9.4 Auditory Cortical Interactions with the Superior Temporal Sulcus Mediates Face/Voice Integration

The finding that there are integrative responses in presumptive unimodal regions such as the auditory cortex does not preclude a role for association cortical areas. The face-specific visual influence on the lateral belt auditory cortex begs the question as to its anatomical source and the likely possibilities for such a source include the superior temporal sulcus (STS), the prefrontal cortex, and the amygdala – regions which have abundant face-sensitive neurons. The STS is likely to be a prominent source for the following reasons. First, there are reciprocal connections between the STS and the lateral belt and other parts of auditory cortex (Barnes and Pandya, 1992; Seltzer and Pandya, 1994). Second, neurons in the STS are sensitive to both faces and biological motion (Harries and Perrett, 1991; Oram and Perrett, 1994). Finally, the STS is known to be multisensory (Barraclough et al., 2005; Benevento et al., 1977; Bruce et al., 1981; Chandrasekaran and Ghazanfar, 2009; Schroeder and Foxe, 2002).

One mechanism for establishing whether auditory cortex and the STS interact at the functional level is to measure their temporal correlations as a function of stimulus condition. Concurrent recordings of LFPs and spiking activity in the lateral belt of auditory cortex and the upper bank of the STS revealed that functional interactions, in the form of gamma band correlations, between these two regions increased in strength during presentations of faces and voices together relative to the unimodal conditions (Ghazanfar et al., 2008) (Fig. 9.4a). Furthermore, these interactions were not solely modulations of response strength, as phase relationships were significantly less variable (tighter) in the multisensory conditions (Fig. 9.4b).

The influence of the STS on auditory cortex was not merely on its gamma oscillations. Spiking activity seems to be *modulated*, but not 'driven', by an ongoing

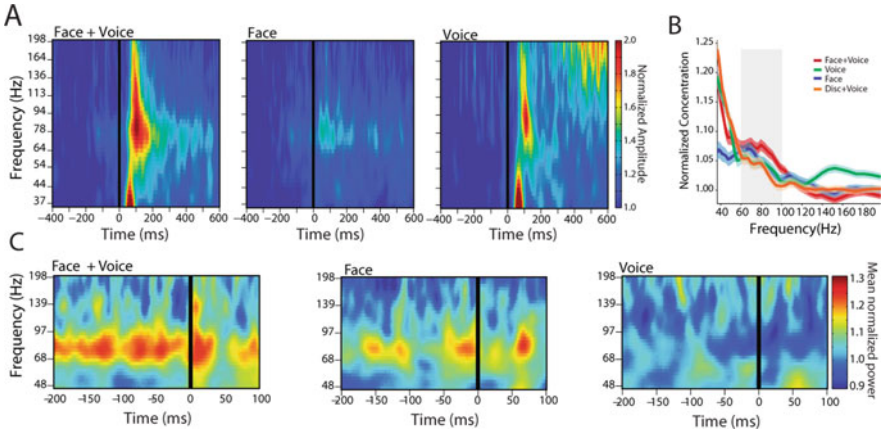

Fig. 9.4 (**a**) Time–frequency plots (cross-spectrograms) illustrate the modulation of functional interactions (as a function of stimulus condition) between the lateral belt auditory cortex and the STS for a population of cortical sites. *x*-axes depict the time in milliseconds as a function of onset of the auditory signal (*solid black line*). *y*-axes depict the frequency of the oscillations in Hz. *Color bar* indicates the amplitude of these signals normalized by the baseline mean. (**b**) Population phase concentration from 0 to 300 ms after voice onset. *x*-axes depict frequency in Hz. *y*-axes depict the average normalized phase concentration. *Shaded regions* denote the SEM across all electrode pairs and calls. All values are normalized by the baseline mean for different frequency bands. *Right panel* shows the phase concentration across all calls and electrode pairs in the gamma band for the four conditions. (**c**) Spike-field cross-spectrogram illustrates the relationship between the spiking activity of auditory cortical neurons and the STS local field potential across the population of cortical sites. *x*-axes depict time in milliseconds as a function of the onset of the multisensory response in the auditory neuron (*solid black line*). *y*-axes depict the frequency in Hz. *Color bar* denotes the cross-spectral power normalized by the baseline mean for different frequencies

activity arising from the STS. Three lines of evidence suggest this scenario. First, visual influences on single neurons were most robust when in the form of dynamic faces and were only apparent when neurons had a significant response to a vocalization (i.e., there were no overt responses to faces alone). Second, these integrative responses were often 'face specific' and had a wide distribution of latencies, which suggested that the face signal was an ongoing signal that influenced auditory responses (Ghazanfar et al., 2008). Finally, this hypothesis for an ongoing signal is supported by the sustained gamma band activity between auditory cortex and STS and by a spike-field coherence analysis of the relationship between auditory cortical spiking activity and gamma oscillations from the STS (Ghazanfar et al., 2008) (Fig. 9.4c).

Both the auditory cortex and the STS have multiple bands of oscillatory activity generated in response to stimuli that may mediate different functions (Chandrasekaran and Ghazanfar, 2009; Lakatos et al., 2005). Thus, interactions between the auditory cortex and the STS are not limited to spiking activity and high-frequency gamma oscillations. Below 20 Hz, and in response to naturalistic audiovisual stimuli, there are directed interactions from auditory cortex to STS, while

above 20 Hz (but below the gamma range), there are directed interactions from STS to auditory cortex (Kayser and Logothetis, 2009). Given that different frequency bands in the STS integrate faces and voices in distinct ways (Chandrasekaran and Ghazanfar, 2009), it's possible that these lower frequency interactions between the STS and the auditory cortex also represent distinct multisensory processing channels.

Although I have focused on the interactions between the auditory cortex and the STS, it is likely that other sources, like the amygdala and the prefrontal cortex, also influence face/voice integration in the auditory cortex. Similar to the STS, these regions are connected to auditory cortex and have an abundance of neurons sensitive to faces and a smaller population sensitive to voices (Gothard et al., 2007; Kuraoka and Nakamura, 2007; Sugihara et al., 2006; Romanski et al., 2005). Indeed, when rhesus monkeys were presented with movies of familiar monkeys vocalizing, approximately half of the neurons recorded in the ventrolateral prefrontal cortex were bimodal in the sense that they responded to both unimodal auditory and visual stimuli or responded differently to bimodal stimuli than to either unimodal stimulus (Sugihara et al., 2006). As in the STS and auditory cortex, prefrontal neurons exhibited enhancement or suppression, and, like the STS but unlike the auditory cortex, suppression (73% of neurons) was more commonly observed than enhancement (27% of neurons).

9.5 Beyond Visual Influences in Auditory Cortex

The auditory cortex is also responsive to other modalities besides vision and these other modalities could also be important for vocal communication. For example, both humans and monkeys tend to look at the eyes more than mouth when viewing vocalizing conspecifics (Ghazanfar et al., 2006; Klin et al., 2002; Vatikiotis-Bateson et al., 1998). When they do fixate on the mouth, it is highly correlated with the onset of mouth movement (Ghazanfar et al., 2006; Lansing and McConkie, 2003). Surprisingly, auditory cortex is likely sensitive to such eye movement patterns: activity in both primary auditory cortex and belt areas is influenced by eye position. For example, when the spatial tuning of primary auditory cortical neurons is measured with the eyes gazing in different directions, ~30% of the neurons are affected by the position of the eyes (Werner-Reiss et al., 2003). Similarly, when LFP-derived current-source density activity was measured from auditory cortex (both primary auditory cortex and caudal belt regions), eye position significantly modulated auditory-evoked amplitude in about 80% of sites (Fu et al., 2004). These eye-position (proprioceptive) effects occurred mainly in the upper cortical layers, suggesting that the signal is fed-back from another cortical area. A possible source includes the frontal eye field (FEF) located in the frontal lobes, the medial portion of which generates relatively long saccades (Robinson and Fuchs, 1969), is interconnected with both the STS (Schall et al., 1995; Seltzer and Pandya, 1989) and multiple regions of the auditory cortex (Hackett et al., 1999; Schall et al., 1995; Romanski et al., 1999).

The auditory cortex is also sensitive to tactile inputs. Numerous lines of both physiological and anatomical evidence demonstrate that at least some regions of the auditory cortex respond to touch as well as sound (Fu et al., 2003; Hackett et al., 2007a, b; Kayser et al., 2005; Lakatos et al., 2007; Schroeder and Foxe, 2002; Smiley et al., 2007). How might tactile signals be involved in vocal communication? Oddly enough, kinesthetic feedback from one's own speech movements also integrates with heard speech (Sams et al., 2005). More directly, if a robotic device is used to artificially deform the facial skin of subjects in a way that mimics the deformation seen during speech production, then subjects actually hear speech differently (Ito et al., 2009). Surprisingly, there is a systematic perceptual variation with speech-like patterns of skin deformation that implicate a robust somatosensory influence on auditory processes under normal conditions (Ito et al., 2009). While the substrates for these somatosensory–auditory effects have not been explored, interactions between the somatosensory system and the auditory cortex seem like a likely source for the phenomena described above for the following reasons. First, many auditory cortical fields respond to, or are modulated by, tactile inputs (Fu et al., 2003; Kayser et al., 2005; Schroeder et al., 2001). Second, there are intercortical connections between somatosensory areas and the auditory cortex (Cappe and Barone, 2005; de la Mothe et al., 2006; Smiley et al., 2007). Third, auditory area CM, where many auditory-tactile responses seem to converge, is directly connected to somatosensory areas in the retroinsular cortex and the granular insula (de la Mothe et al., 2006; Smiley et al., 2006). Finally, the tactile receptive fields of neurons in auditory cortical area CM are confined to the upper body, primarily the face and neck regions (areas consisting of, or covering, the vocal tract) (Fu et al., 2003) and the primary somatosensory cortical (area 3b) representation for the tongue (a vocal tract articulator) projects to auditory areas in the lower bank of the lateral sulcus (Iyengar et al., 2007). All of these facts lend further credibility to the putative role of somatosensory–auditory interactions during vocal production and perception.

9.6 The Development of Multisensory Systems for Communication

While it appears that monkeys and humans share numerous behavioral and neural phenotypes related to multisensory integration of communication signals, how these systems emerge may not be identical. Given that monkeys and humans develop at different rates, it is important to know how this difference in developmental timing (or *heterochrony*) might influence the behavior and neural circuitry underlying multisensory communication. One line of investigation suggests that an interaction between developmental timing (*heterochrony*) and social experience may shape the neural circuits underlying both human and primate vocal communication (Lewkowicz and Ghazanfar, 2006; Lewkowicz et al., 2008; Zangehenpour et al., 2008).

The rate of neural development in Old World monkeys is faster than in humans and, as a result, they are neurologically precocial relative to humans. For example, in terms of overall brain size at birth, Old World monkeys are among the most precocial of all mammals (Sacher and Staffeldt, 1974), possessing ∼65% of their brain size at birth compared to only ∼25% for human infants (Malkova et al., 2006; Sacher and Staffeldt, 1974). Second, fiber pathways in the developing monkey brain are more heavily myelinated than in the human brain at the same postnatal age (Gibson, 1991), suggesting that postnatal myelination in the rhesus monkey brain is about three to four times faster than in the human brain (Gibson, 1991; Malkova et al., 2006). All sensorimotor tracts are heavily myelinated by 2–3 months after birth in rhesus monkeys, but not until 8–12 months after birth in human infants. Finally, at the behavioral level, the differential patterns of brain growth in the two species lead to differential timing in the emergence of species-specific motor, socio-emotional, and cognitive abilities (Antinucci, 1989; Konner, 1991).

The differences in the timing of neural and behavioral development across different primate species raise the possibility that the development of intersensory integration may be different in monkeys relative to humans. The slow rate of neural development in human infants (relative to monkeys) may actually be advantageous because their altricial brains may provide them with greater functional plasticity and better correspondence with their postnatal environment. As a consequence, however, this may make human infants initially more sensitive to a broader range of sensory stimulation and to the relations among multisensory inputs. This theoretical observation has received empirical support from studies showing that infants go through a process of 'perceptual narrowing' in their processing of unisensory as well as multisensory information; that is, where initially they exhibit broad sensory tuning, they later exhibit narrower tuning. For example, 4- to 6-month-old human infants can match rhesus monkey faces and voices, but 8- to 10-month-old infants no longer do so (Lewkowicz and Ghazanfar, 2006). These findings suggest that as human infants acquire increasingly greater experience with conspecific human faces and vocalizations – but none with heterospecific faces and vocalizations – their sensory tuning (and their neural systems) narrows to match their early experience.

If a relatively immature state of neural development leaves a developing human infant more 'open' to the effects of early sensory experience, then it stands to reason that the more advanced state of neural development in monkeys might result in a different outcome. In support of this, a study of infant vervet monkeys that was identical in design to the human infant study of cross-species multisensory matching (Lewkowicz and Ghazanfar, 2006) revealed that, unlike human infants, they exhibit no evidence of perceptual narrowing (Zangehenpour et al., 2008). That is, the infant vervet monkeys could match faces and voices of rhesus monkeys despite the fact that they had no prior experience with macaque monkeys and that they continued to do so well beyond the ages where such matching ability declines in human infants (Zangehenpour et al., 2008). The reason for this lack of perceptual narrowing may lie in the precocial neurological development of this Old World monkey species.

These comparative developmental data reveal that while monkeys and humans may appear to share similarities at the behavioral and neural levels, their different

developmental trajectories are likely to reveal important differences. It is important to keep this in mind when making claims about homologies at either of these levels.

9.7 Conclusion

Communication is, by default, a multisensory phenomenon. This is evident in the automatic integration of the senses during vocal perception in humans and monkeys, the evidence of such integration early in development, and most importantly, by the organization of the neocortex. The overwhelming evidence from the studies reviewed here, and numerous other studies from different domains of neuroscience, all converge on the idea that the neocortex is fundamentally multisensory (Ghazanfar and Schroeder, 2006). It is not confined to a few 'sensu comune' in the association cortices. It is all over. This does not mean, however, that every cortical area is uniformly multisensory, but rather that cortical areas maybe weighted differently by 'extra'-modal inputs depending on the task at hand and its context.

Acknowledgments The author gratefully acknowledges the scientific contributions and numerous discussions with the following people: Chand Chandrasekaran, Kari Hoffman, David Lewkowicz, Joost Maier, and Hjalmar Turesson. This work was supported by NIH R01NS054898 and NSF BCS-0547760 CAREER Award.

References

Adachi I, Kuwahata H, Fujita K, Tomonaga M, Matsuzawa T (2006) Japanese macaques form a cross-modal representation of their own species in their first year of life. Primates 47: 350–354

Antinucci F (1989) Systematic comparison of early sensorimotor development. In: Antinucci F (ed) Cognitive structure and development in nonhuman primates.: Lawrence Erlbaum Associates, Hillsdale, NJ, pp 67–85

Barnes CL, Pandya DN (1992) Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus-monkey. J Comp Neurol 318:222–244

Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cogn Neurosci 17:377–391

Benevento LA, Fallon J, Davis BJ, Rezak M (1977) Auditory-visual interactions in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. Exp Neurol 57:849–872

Bernstein LE, Auer ET, Takayanagi S (2004) Auditory speech detection in noise enhanced by lipreading. Speech Commun 44:5–18

Besle J, Fort A, Delpuech C, Giard MH (2004) Bimodal speech: early suppressive visual effects in human auditory cortex. Eur J Neurosci 20:2225–2234

Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb Cortex 17:2172–2189

Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 46:369–384

Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. Eur J Neurosci 22:2886–2902

Chandrasekaran C, Ghazanfar AA (2009) Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. J Neurophysiol 101:773–788
Cheney DL, Seyfarth RM (1982) How vervet monkeys perceive their grunts – field playback experiments. Animal Behav 30:739–751
de la Mothe LA, Blumell S, Kajikawa Y, Hackett TA (2006) Cortical connections of the auditory cortex in marmoset monkeys: Core and medial belt regions. J Comp Neurol 496:27–71
Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron 57:11–23
Ettlinger G, Wilson WA (1990) Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms. Behav Brain Res 40:169–192
Evans TA, Howell S, Westergaard GC (2005) Auditory-visual cross-modal perception of communicative stimuli in tufted capuchin monkeys (Cebus apella). J Exp Psychol-Anim Behav Proc 31:399–406
Fitch WT (1997) Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. J Acoust Soc Am 102:1213–1222
Fitch WT, Hauser MD (1995) Vocal production in nonhuman-primates - acoustics, physiology, and functional constraints on honest advertisement. Am J Primatol 37:191–219
Fu KMG, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE (2003) Auditory cortical neurons respond to somatosensory stimulation. J Neurosci 23: 7510–7515
Fu KMG, Shah AS, O'Connell MN, McGinnis T, Eckholdt H, Lakatos P, Smiley J, Schroeder CE (2004) Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. J Neurophysiol 92:3522–3531
Ghazanfar AA, Logothetis NK (2003) Facial expressions linked to monkey calls. Nature 423: 937–938
Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278–285
Ghazanfar AA, Rendall D (2008) Evolution of human vocal production. Curr Biol 18:R457–R460
Ghazanfar AA, Nielsen K, Logothetis NK (2006) Eye movements of monkeys viewing vocalizing conspecifics. Cognition 101:515–529
Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. J Neurosci 28:4457–4469
Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J Neurosci 25:5004–5012
Ghazanfar AA, Turesson HK, Maier JX, van Dinther R, Patterson RD, Logothetis NK (2007) Vocal tract resonances as indexical cues in rhesus monkeys. Curr Biol 17:425–430
Gibson KR (1991) Myelination and behavioral development: A comparative perspective on questions of neoteny, altriciality and intelligence. In: Gibson KR, Petersen AC (eds) Brain maturation and cognitive development: comparative and cross-cultural perspective. Aldine de Gruyter, New York, pp 29–63
Gogate LJ, Walker-Andrews AS, Bahrick LE (2001) The intersensory origins of word comprehension: an ecological-dynamic systems view. Develop Sci 4:1–18
Gothard KM, Battaglia FP, Erickson CA, Spitler KM, Amaral DG (2007) Neural responses to facial expression and face identity in the monkey amygdala. J Neurophysiol 97:1671–1683
Hackett TA, Stepniewska I, Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. Brain Res 817:45–58
Hackett TA, De La Mothe LA, Ulbert I, Karmos G, Smiley J, Schroeder CE (2007a) Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. J Comp Neurol 502:924–952
Hackett TA, Smiley JF, Ulbert I, Karmos G, Lakatos P, de la Mothe LA, Schroeder CE (2007b) Sources of somatosensory input to the caudal belt areas of auditory cortex. Perception 36: 1419–1430

Harries MH, Perrett DI (1991) Visual processing of faces in temporal cortex - physiological evidence for a modular organization and possible anatomical correlates. J Cogn Neurosci 3:9–24
Hauser MD, Ybarra MS (1994) The role of lip configuration in monkey vocalizations - experiments using xylocaine as a nerve block. Brain Lang 46:232–244
Hauser MD, Evans CS, Marler P (1993) The role of articulation in the production of rhesus-monkey, Macaca-Mulatta, vocalizations. Anim Behav 45:423–433
Ito T, Tiede M, Ostry DJ (2009) Somatosensory function in speech perception. Proc Natl Acad Sci U S A 106:1245–1248
Iyengar S, Qi H, Jain N, Kaas JH (2007) Cortical and thalamic connections of the representations of the teeth and tongue in somatosensory cortex of new world monkeys. J Comp Neurol 501: 95–120
Izumi A, Kojima S (2004) Matching vocalizations to vocalizing faces in a chimpanzee (Pan troglodytes). Anim Cogn 7:179–184
Jiang JT, Alwan A, Keating PA, Auer ET, Bernstein LE (2002) On the relationship between face movements, tongue movements, and speech acoustics. Eurasip J Appl Sig Proc 2002: 1174–1188
Jordan KE, Brannon EM, Logothetis NK, Ghazanfar AA (2005) Monkeys match the number of voices they hear with the number of faces they see. Curr Biol 15:1034–1038
Kayser C, Logothetis NK (2009) Directed interactions between auditory and superior temporal cortices and their role in sensory integration. Front Integr Neurosci 3:7
Kayser C, Petkov CI, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. Cereb Cortex 18:1560–1574
Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. Neuron 48:373–384
Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional imaging reveals visual modulation of specific fields in auditory cortex. J Neurosci 27:1824–1835
Klin A, Jones W, Schultz R, Volkmar F, Cohen D (2002) Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Archiv Gen Psychiatry 59:809–816
Konner M (1991) Universals of behavioral development in relation to brain myelination. In: Gibson KR, Petersen AC (eds) Brain maturation and cognitive development: comparative and cross-cultural perspectives. Aldine de Gruyter, New York, pp 181–223
Kuhl PK, Williams KA, Meltzoff AN (1991) Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. J Exp Psychol: Human Percept Perform 17:829–840
Kuraoka K, Nakamura K (2007) Responses of single neurons in monkey amygdala to facial and vocal emotions. J Neurophysiol 97:1379–1387
Lakatos P, Chen C-M, O′Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53:279–292
Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904–1911
Lansing IR, McConkie GW (2003) Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. Percept Psychophys 65: 536–552
Lewkowicz DJ, Ghazanfar AA (2006) The decline of cross-species intersensory perception in human infants. Proc Natl Acad Sci U S A 103:6771–6774
Lewkowicz DJ, Sowinski R, Place S (2008) The decline of cross-species intersensory perception in human infants: underlying mechanisms and its developmental persistence. Brain Res 1242:291–302
Malkova L, Heuer E, Saunders RC (2006) Longitudinal magnetic resonance imaging study of rhesus monkey brain development. Eur J Neurosci 24:3204–3212
McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:229–239

Oram MW, Perrett DI (1994) Responses of anterior superior temporal polysensory (Stpa) neurons to biological motion stimuli. J Cogn Neurosci 6:99–116
Palombit RA, Cheney DL, Seyfarth RM (1999) Male grunts as mediators of social interaction with females in wild chacma baboons (Papio cynocephalus ursinus). Behaviour 136: 221–242
Parr LA (2004) Perceptual biases for multimodal cues in chimpanzee (Pan troglodytes) affect recognition. Anim Cogn 7:171–178
Patterson ML, Werker JF (2003) Two-month-old infants match phonetic information in lips and voice. Develop Sci 6:191–196
Pevsner J (2002) Leonardo da Vinci's contributions to neuroscience. Trends Neurosci 25: 217–220
Robinson DA, Fuchs AF (1969) Eye movements evoked by stimulation of frontal eye fields. J Neurophysiol 32:637–648
Romanski LM, Bates JF, Goldman-Rakic PS (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. J Comp Neurol 403:141–157
Romanski LM, Averbeck BB, Diltz M (2005) Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J Neurophysiol 93:734–747
Rosenblum LD (2005) Primacy of multimodal speech perception. In: Pisoni DB, Remez RE (eds) Handbook of speech perception. Blackwell, Malden, MA, pp 51–78
Rosenblum LD (2008) Speech perception as a multimodal phenomenon. Curr Direct Psychol Sci 17:405–409
Sacher GA, Staffeldt EF (1974) Relation of gestation time to brain weight for placental mammals: implications for the theory of vertebrate growth. Am Naturalist 108:593–615
Sams M, Mottonen R, Sihvonen T (2005) Seeing and hearing others and oneself talk. Cogn Brain Res 23:429–435
Schall JD, Morel A, King DJ, Bullier J (1995) Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. J Neurosci 15: 4464–4487
Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cogn Brain Res 14:187–198
Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC (2001) Somatosensory input to auditory association cortex in the macaque monkey. J Neurophysiol 85:1322–1327
Schwartz J-L, Berthommier F, Savariaux C (2004) Seeing to hear better: evidence for early audio-visual interactions in speech identification. Cognition 93:B69–B78
Seltzer B, Pandya DN (1989) Frontal-lobe connections of the superior temporal sulcus in the rhesus-monkey. J Comp Neurol 281:97–113
Seltzer B, Pandya DN (1994) Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. J Comp Neurol 343: 445–463
Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE (2007) Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. J Comp Neurol 502:894–923
Sugihara T, Diltz MD, Averbeck BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. J Neurosci 26: 11138–11147
van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci US A 102:1181–1186
Vatikiotis-Bateson E, Eigsti IM, Yano S, Munhall KG (1998) Eye movement of perceivers during audiovisual speech perception. Percept Psychophys 60:926–940
Werner-Reiss U, Kelly KA, Trause AS, Underhill AM, Groh JM (2003) Eye position affects activity in primary auditory cortex of primates. Curr Biol 13:554–562
Yehia H, Rubin P, Vatikiotis-Bateson E (1998) Quantitative association of vocal-tract and facial behavior. Speech Commun 26:23–43

Yehia HC, Kuratate T, Vatikiotis-Bateson E (2002) Linking facial animation, head motion and speech acoustics. J Phonet 30:555–568

Zangehenpour S, Ghazanfar AA, Lewkowicz DJ, Zatorre RJ (2008) Heterochrony and cross-species intersensory matching by infant vervet monkeys. PLoS ONE 4:e4302