The multisensory roles for auditory cortex in primate vocal communication

Asif A. Ghazanfar

Neuroscience Institute, Departments of Psychology and Ecology & Evolutionary Biology, Princeton University, Princeton, NJ 08540, USA

ARTICLE INFO

ABSTRACT

Primate vocal communication is a fundamentally multisensory behavior and this will be reflected in the different roles brain regions play in mediating it. Auditory cortex is illustrative, being influenced, I will argue, by the visual, somatosensory, proprioceptive and motor modalities during vocal communication. It is my intention that the data reviewed here suggest that investigating auditory cortex through the lens of a specific behavior may lead to a much clearer picture of its functions and dynamic organization. One possibility is that, beyond its tonotopic and cytoarchitectural organization, the auditory cortex may be organized according to ethologically-relevant actions. Such action-specific representations would be overlaid on top of traditional mapping schemes and would help mediate motor and multisensory processes related to a particular type of behavior.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

It is widely accepted that human speech is fundamentally a multisensory behavior, with face-to-face communication perceived through both the visual and auditory channels. Such multisensory speech perception is evident even at the earliest stages of human cognitive development (Gogate et al., 2001; Patterson and Werker, 2003); its integration across the two modalities is ubiquitous and automatic (McGurk and MacDonald, 1976); and at the neural level, audiovisual speech integration occurs at the 'earliest' stages of cortical processing (Ghazanfar and Schroeder, 2006). Indeed, there are strong arguments which suggest that multisensory speech is the *primary* mode of speech perception and is not a capacity that is "piggybacked" on to auditory speech perception (Rosenblum, 2005). This implies that the perceptual mechanisms, neurophysiology and evolution of speech perception are based on primitives which are not tied to a single sensory modality (Romanski and Ghazanfar, in press). The essence of these ideas is shared by many investigators in the domain of perception (Fowler, 2004; Liberman and Mattingly, 1985; Meltzoff and Moore, 1997), but has only caught on belatedly for those of us who study the auditory cortex.

The auditory cortex of primates consists of numerous fields (Hackett et al., 1998; Petkov et al., 2006), which in turn are connected to numerous auditory-related areas in the frontal and temporal lobes (Romanski and Ghazanfar, in press). At least a subset of these fields appears to be homologous across monkeys and

great apes (including humans) (Hackett et al., 2001). These fields are delineated largely by their tonotopic organization and anatomical criteria. The reasons for why there are so many areas are not known, and how each of them, together or separately, relate to behavior is also somewhat of a mystery. That they must be involved in multiple auditory-related behaviors is a given. The fundamental question is thus: how do these multiple auditory areas mediate specific behaviors through their interactions with each other and with other sensory and motor systems?

The question posed above has an underlying assumption that the different auditory cortical fields do not each have a specific function that is used in all types of auditory-related behavior. Rather, their roles (or weights) as nodes in a larger network change according to whatever specific behavior is being mediated. In this review, I focus on one behavior, vocal communication, to illustrate how multiple fields of auditory cortex in non-human primates (hereafter, *primates*) may play a role in the perception and production of this multisensory behavior (see Kayser et al, this volume, for a general overview of auditory cortical organization and its relationship to visual and somatosensory inputs). I will begin by briefly presenting evidence that primates do indeed link visual and auditory communication signals, and then describe how such perception may be mediated by the auditory cortex through its connections with association areas. I will then speculate as to how proprioceptive, somatosensory and motor inputs into auditory cortex also play a role in vocal communication. Finally, I will conclude with the suggestion that the use of cytoarchitecture and tonotopy represent one way of defining auditory cortical organization, and that there may be behavior-specific functional organizations that do not neatly fall within the delineation of

E-mail address: asifg@princeton.edu

different auditory cortical fields using these methods. Identifying specific behaviors *a priori* is the key to illuminating such putative organizational schemes.

2. Monkeys link facial expressions to vocal expressions

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

Given that both humans and other extant primates use both facial and vocal expressions as communication signals, it is perhaps not surprising that many primates other than humans recognize the correspondence between the visual and auditory components of vocal signals. Macaque monkeys (*Macaca mulatta*), capuchins (*Cebus apella*) and chimpanzees (*Pan troglodytes*) all recognize auditory-visual correspondences between their various vocalizations (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). Finally, macaque monkeys use formants (i.e., vocal tract res-

onances) as acoustic cues to assess age-related body size differences among conspecifics (Ghazanfar et al., 2007). 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).

3. Dynamic faces modulate voice processing in auditory cortex

Traditionally, 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., 2008, 2005; Kayser et al., 2008, 2007; Schroeder and Foxe, 2002). Here, we focus on those auditory cortical studies investigating face/voice integration specifically.

Recordings from both primary and lateral belt auditory cortex reveal that responses to the voice are influenced by the presence of a dynamic face (Ghazanfar et al., 2008, 2005). 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 primary auditory cortex and 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. 2A). By 'integra-
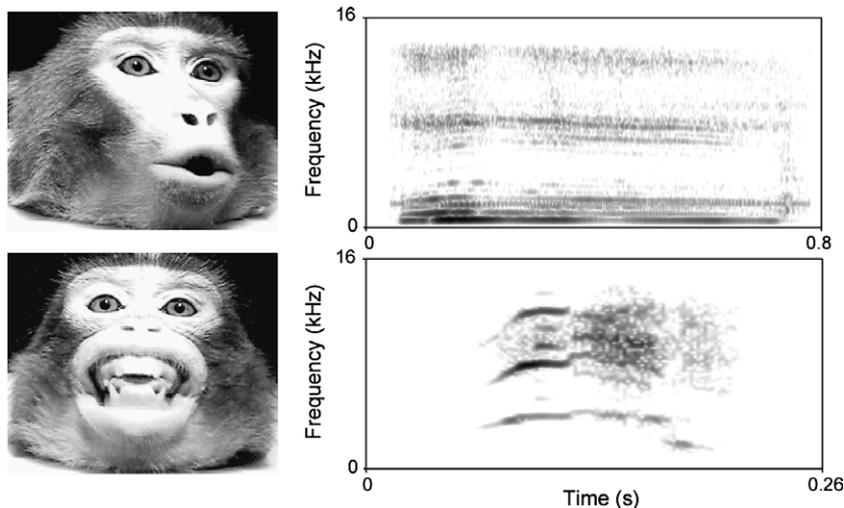
Fig. 1. Exemplars of the facial expressions produced concomitantly with vocalizations. (A) Rhesus monkey coo and scream calls taken at the midpoint of the expressions with their corresponding spectrograms.
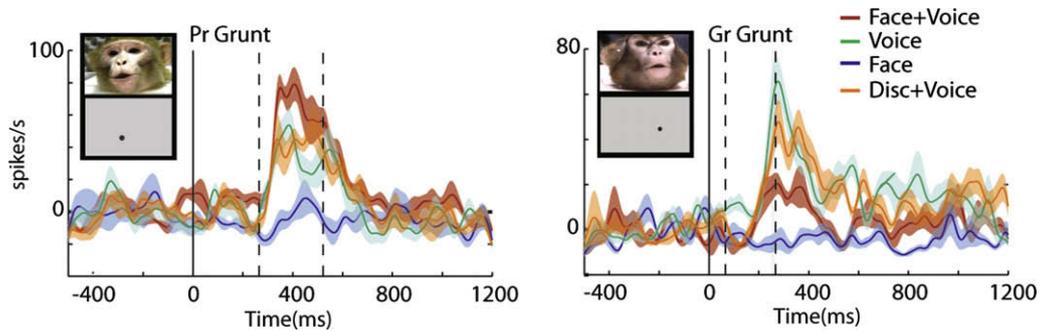
Fig. 2. Single neuron examples of multisensory integration of Face + Voice stimuli compared with Disk + Voice stimuli in the middle lateral (ML) belt area. The left panel shows an enhanced response when voices are coupled with faces, but no similar modulation when coupled with disks. The right panel 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. (See Ghazanfar et al., 2008 for details).

tion', I simply mean that bimodal stimuli elicit significantly enhanced or suppressed responses relative to the best (strongest) response elicited by unimodal stimuli.

The specificity of face/voice integrative responses was tested by replacing the dynamic faces 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 responses elicited with real monkey faces (Ghazanfar et al., 2008, 2005) (Fig. 2). 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). This is perhaps not surprising given that the lateral belt is well known for its responsiveness and, to some degree, selectivity, to vocalizations and other complex stimuli (Rauschecker et al., 1995; Recanzone, 2008). (See Ghazanfar and Hauser, 1999, 2001 for review of vocal responses in auditory cortex.)

Unexpectedly, grunt vocalizations were over-represented relative to coos in terms of enhanced multisensory LFP responses (Ghazanfar et al., 2005). As coos and grunts are both produced frequently in a variety of affiliative contexts and are broadband spectrally, the differential representation cannot be attributed to auditory 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 towards 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 towards grunt calls may be related to the fact the grunts (relative to coos) are often produced during intimate, one-to-one social interactions.
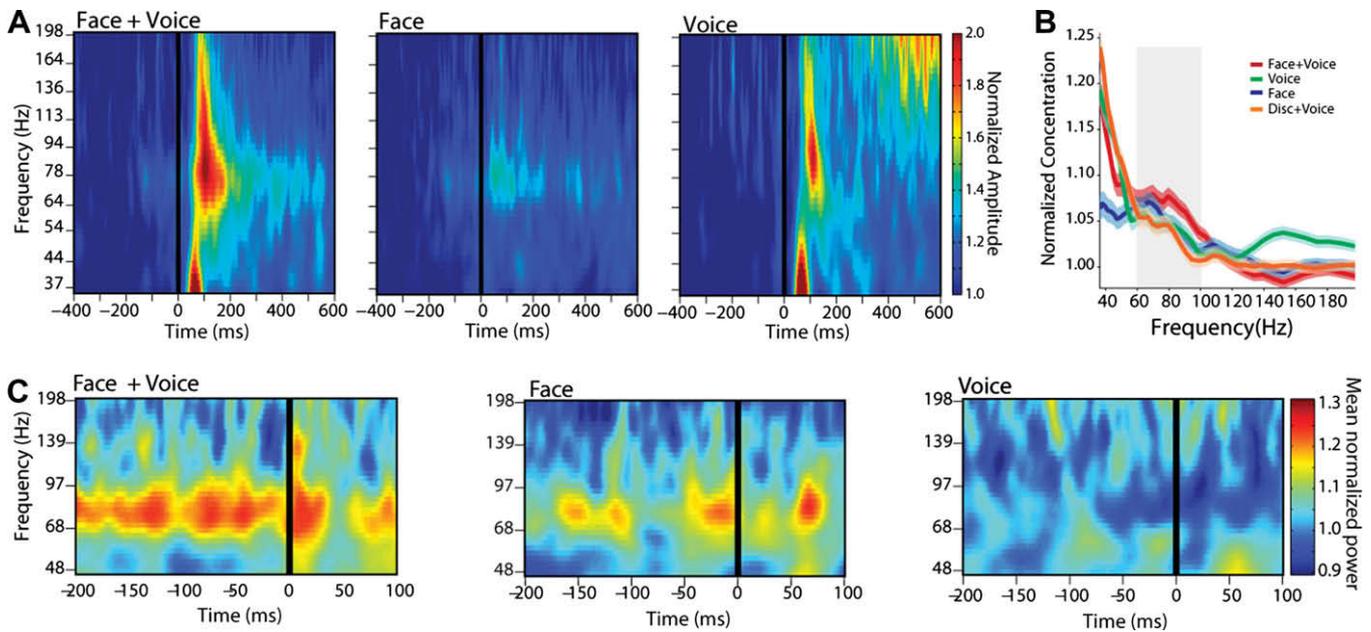
4. Auditory cortical interactions with the superior temporal sulcus mediates face/voice integration

The "face-specific" visual influence on the lateral belt auditory cortex begs the question as to its anatomical source. Although there are multiple possible sources of visual input to auditory cortex (Ghazanfar and Schroeder, 2006), the superior temporal sulcus (STS) is likely to be a prominent one, particularly for integrating faces and voices, 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 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 (>30 Hz) correlations, between these two regions increased in strength during presentations of faces and voices together relative to the unimodal or dynamic disk conditions (Ghazanfar et al., 2008) (Fig. 3A). Furthermore, these interactions were not solely modulations of response strength, as phase relationships were significantly less variable (tighter) in the multisensory conditions (Fig. 3B).

The influence of the STS on auditory cortex was not merely on its gamma oscillations. Spiking activity in auditory cortex seems to be *modulated*, but not 'driven', by on-going 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. 3C).

Both the auditory cortex and the STS have multiple bands of oscillatory activity generated in responses to communication signals 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 (in press)). Given that different frequency bands in the STS integrate faces and voices in distinct ways (Chandrasekaran and Ghazanfar, 2009),

Fig. 3. (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 0–300 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 (See Ghazanfar et al., 2008 for details).

it's possible that these lower frequency interactions between the STS and auditory cortex also represent distinct multisensory processing channels.

Two things should be noted here. The first is that functional interactions between STS and auditory cortex are not likely to occur solely during the presentation of faces with voices. Other congruent, behaviorally-salient audiovisual events such as looming signals (Cappe et al., 2009; Gordon and Rosenblum, 2005; Maier et al., 2004) or other temporally coincident signals may elicit similar functional interactions (Maier et al., 2008; Noesselt et al., 2007). The second is that there are other areas that, consistent with their connectivity and response properties (e.g., sensitivity to faces and voices), could also (and very likely) have a visual influence on auditory cortex. These include the ventrolateral prefrontal cortex (Romanski et al., 2005; Sugihara et al., 2006) and the amygdala (Gothard et al., 2007; Kuraoka and Nakamura, 2007). It is not known whether STS, for instance, plays a more influential role than these two other 'face sensitive' areas. Indeed, it may be that all three play very different roles in face/voice integration. What is missing is a direct link between multisensory behaviors and neural activity – that is only way to assess the true contributions of these regions, along with auditory cortex, in vocal behavior.
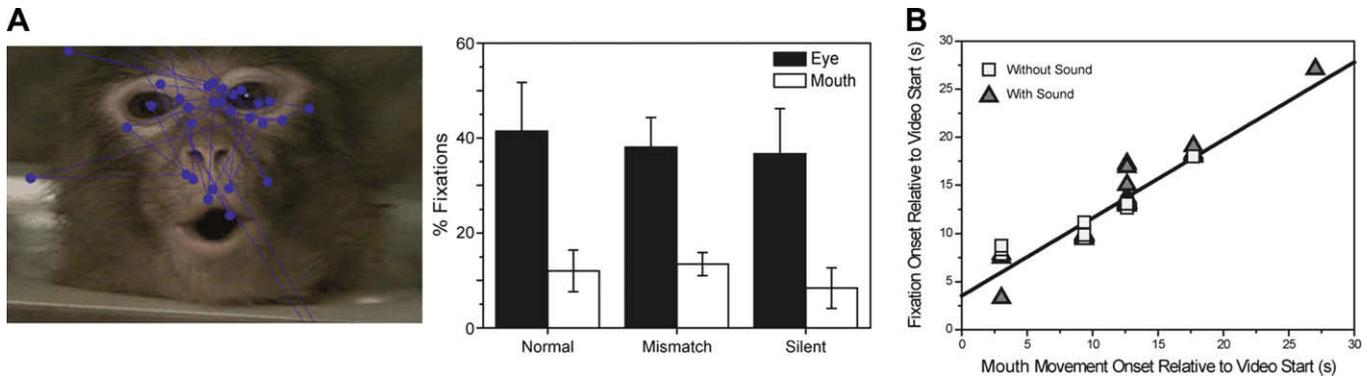
5. Viewing vocalizing conspecifics

Humans and other primates readily link facial expressions with appropriate, congruent vocal expressions. What cues they use to make such matches are not known. One method for investigating such behavioral strategies is the measurement of eye movement patterns. When human subjects are given *no* task or instruction regarding what acoustic cues to attend, they will consistently look at the eye region more than the mouth when viewing videos of human speakers (Klin et al., 2002). Macaque monkeys exhibit the exact same strategy. The eye movement patterns of monkeys viewing conspecifics producing vocalizations reveal that monkeys spend most of their time inspecting the eye region relative to the mouth (Ghazanfar et al., 2006) (Fig. 4A). When they did fixate on the mouth, it was highly correlated with the onset of mouth movements (Fig. 4B). This, too, was highly reminiscent of human strategies: subjects asked to identify words increased their fixations onto the mouth region with the onset of facial motion (Lansing and McConkie, 2003).

Somewhat surprisingly, activity in both primary auditory cortex and belt areas is influenced by eye position. 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 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; Romanski et al., 1999; Schall et al., 1995).

It does not take a huge stretch of the imagination to link these auditory cortical processes to the oculomotor strategy for looking at vocalizing faces. A dynamic, vocalizing face is a complex sequence of sensory events, but one that elicits fairly stereotypical

Fig. 4. (A) The average fixation on the eye region versus the mouth region across three subjects while viewing a 30-s video of vocalizing conspecific. The audio track had no influence on the proportion of fixations falling onto the mouth or the eye region. Error bars represent SEM. (B) We also find that when monkeys do saccade to the mouth region, it is tightly correlated with the onset of mouth movements ($r = 0.997$, $p < 0.00001$) (See Ghazanfar et al., 2006 for details).

eye movements: we and other primates fixate on the eyes but then saccade to mouth when it moves before saccading back to the eyes. Is there a simple scenario that could link the proprioceptive eye-position effects in the auditory cortex with its face/voice integrative properties (Ghazanfar and Chandrasekaran, 2007)? Reframing (ever so slightly) the hypothesis of Schroeder and colleagues (Lakatos et al., 2007; Schroeder et al., 2008), one possibility is that the fixations at the onset of mouth movements sends a signal to the auditory cortex which resets the phase of an on-going oscillation. This proprioceptive signal thus primes the auditory cortex to amplify or suppress (depending on the timing) of a subsequent auditory signal originating from the mouth. Given that mouth movements precede the voiced components of both human (Abry et al., 1996) and monkey vocalizations (Chandrasekaran and Ghazanfar, 2009; Ghazanfar et al., 2005), the temporal order of visual to proprioceptive to auditory signals is consistent with this idea. This hypothesis is also supported (though indirectly) by the finding that sign of face/voice integration in the auditory cortex and the STS is influenced by the timing of mouth movements relative to the onset of the voice (Chandrasekaran and Ghazanfar, 2009; Ghazanfar et al., 2005).

## 6. Somatosensory feedback during vocal communication

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). Yet, the sense of touch is not something we normally associate with vocal communication. It can, however, influence what we hear under certain circumstances. For example, 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).

The somatosensory system's influence on the auditory system may also occur during vocal learning. When a mechanical load is applied to the jaw, causing a slight protrusion, as subjects repeat words ('saw', 'say', 'sass' and 'sane') it can alter somatosensory feedback without changing the acoustics of the words (Tremblay et al., 2003). Measuring adaptation in the jaw trajectory after many trials revealed that subjects learn to change their jaw trajectories

so that they are similar to the pre-load trajectory – despite not hearing anything different. This strongly implicates a role for somatosensory feedback that parallels the role for auditory feedback in guiding vocal production (Jones and Munhall, 2003, 2005). Indeed, the very same learning effects are observed with deaf subjects when they turn their hearing aids off (Nasir and Ostry, 2008).

While the substrates for these somatosensory-auditory effects have not been explored, interactions between the somatosensory system and the auditory cortex seem to be 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., 2007). Oddly enough, a parallel influence of audition on somatosensory areas has also been reported: neurons in the "somatosensory" insula readily and selectively respond to vocalizations (Beiser, 1998; Remedios et al., 2009). 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) (Fig. 5) 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.

Like humans, other primates also adjust their vocal output according to what they hear. For example, macaques, marmosets (*Callithrix jacchus*), and cotton-top tamarins (*Saguinus oedipus*) adjust the loudness, timing and acoustic structure of their vocalizations depending background noise levels and patterns (Brumm et al., 2004; Egnor and Hauser, 2006; Sinnott et al., 1975); (Egnor et al., 2006, 2007). The specific number of syllables and temporal modulations in heard conspecific calls can also differentially trigger vocal production in tamarins (Ghazanfar et al., 2001, 2002). Thus, auditory feedback is also very important for non-human primates and altering such feedback can influence neurons in the auditory cortex (Eliades and Wang, 2008). At this time, however, no experiments have investigated whether somatosensory feedback plays a role in influencing vocal feedback. The neurophysiological and neuroanatomical data described above suggest that it is not unreasonable to think that it does.

A

B
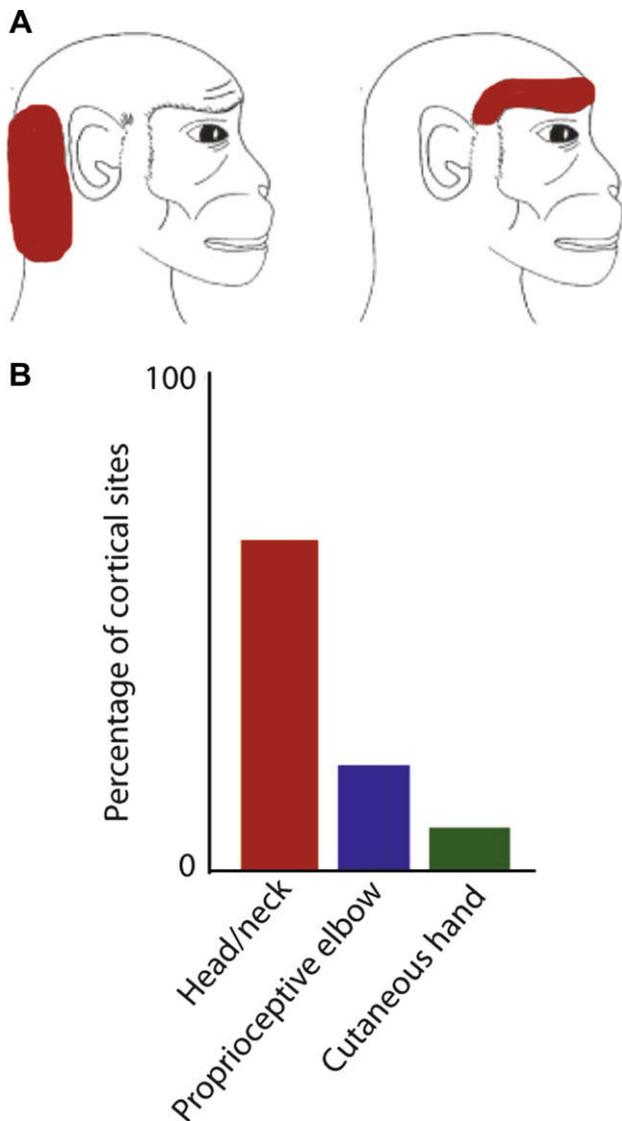
Fig. 5. (A) Examples of cutaneous receptive fields of neurons recorded in auditory cortical area CM. (B) Proportions of different types of somatosensory receptive fields recorded from auditory cortical area CM. x-axis is percentage of multiunit sites (both figures redrawn from Fu et al., 2003).

7. Is auditory cortex organized by action-specific representations?

The putative neural processes underlying multisensory vocal communication in primates calls to mind what the philosopher, Andy Clark, refers to as "action-oriented" representations (Clark, 1997). In generic terms, action-oriented representations simultaneously describe aspects of the world and prescribe possible actions. They are poised between pure control (motor) structures and passive (sensory) representations of the external world. For neural representations of primate vocal communication, this suggests that the laryngeal, articulatory and respiratory movements during vocalizations are inseparable from the visual, auditory and somatosensory processes that accompany vocal perception. This idea seems to fit well with the data reviewed above and suggests an alternative way of thinking about auditory cortical organization.

Typically, we think of auditory cortex as a set of very discrete fields, most of which can be defined by a tonotopic map. These physiological maps often correspond to cytoarchitectural and hodological signatures as well. It is possible, however, that this is just

one of many possible schemes for auditory cortical organization (albeit an important one). One alternative is that different behaviors, such as multisensory vocal communication, each have their own organizational scheme super-imposed on these tonotopic maps, but not necessarily in one-to-one fashion. This is *almost* pure speculation at this point, but the main reason for thinking it a possibility is that many of the multisensory and vocalization-related responses properties for communication have no relationship to the tonotopic maps for, or frequency tuning of neurons in, those regions. For example, there is no reported relationship between the frequency tuning of auditory cortical neurons and the influence of eye position (Fu et al., 2004; Werner-Reiss et al., 2003), somatosensory receptive fields (Fu et al., 2003; Schroeder et al., 2001), or face/voice integration (Ghazanfar et al., 2005; Ghazanfar, unpublished observations). Likewise, the influence of vocal feedback on auditory cortex has no relationship to the underlying frequency tuning of neurons (Eliades and Wang, 2008). At a more global level, somatosensory and visual influences on the auditory cortex seem to take the form of a gradient, extending in the posterior-to-anterior direction, rather than having a discrete influence on particular tonotopically-defined subsets (Kayser et al., 2005, 2007). Finally, the representation for pitch processing, important for vocal recognition, spans the low frequency borders of two core auditory cortical areas (primary auditory cortex and area R) (Bendor and Wang, 2005), violating the often implicit assumption of "single area-single function" rule. The idea that there may be multiple behaviorally-specific auditory cortical organizations is very similar to the one recently put forth regarding the organization of the various somatotopically-defined motor cortical areas (Graziano and Aflalo, 2007).

To summarize, vocal communication is a fundamentally multisensory behavior and this will be reflected in the different roles brain regions play in mediating it. Auditory cortex is illustrative, being influenced by the visual, somatosensory, proprioceptive and motor modalities during vocal communication. In all, I hope that the data reviewed above suggest that investigating auditory cortex through the lens of a *specific behavior* may lead to a much clearer picture of its functions and dynamic organization.

References

Abry, C., Lallouache, M.-T., Cathiard, M.-A., 1996. How can coarticulation models account for speech sensitivity in audio-visual desynchronization? In: Stork, D., Henneke, M. (Eds.), Speechreading by Humans and Machines: Models, Systems and Applications. Springer-Verlag, Berlin, pp. 247–255.

Barnes, C.L., Pandya, D.N., 1992. Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus-monkey. Journal of Comparative Neurology 318, 222–244.

Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., Perrett, D.I., 2005. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. Journal of Cognitive Neuroscience 17, 377–391.

Beiser, A., 1998. Processing of twitter-call fundamental frequencies in insula and auditory cortex of squirrel monkeys. Experimental Brain Research 122, 139–148.

Bendor, D., Wang, X.Q., 2005. The neuronal representation of pitch in primate auditory cortex. Nature 436, 1161–1165.

Benevento, L.A., Fallon, J., Davis, B.J., 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. Experimental Neurology 57, 849–872.

Bernstein, L.E., Auer, E.T., Takayanagi, S., 2004. Auditory speech detection in noise enhanced by lipreading. Speech Communication 44, 5–18.

Besle, J., Fort, A., Delpuech, C., Giard, M.H., 2004. Bimodal speech: early suppressive visual effects in human auditory cortex. European Journal of Neuroscience 20, 2225–2234.

Bizley, J.K., Nodal, F.R., Bajo, V.M., Nelken, I., King, A.J., 2007. Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cerebral Cortex 17, 2172–2189.

Bruce, C., Desimone, R., Gross, C.G., 1981. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. Journal of Neurophysiology 46, 369–384.

Brumm, H., Voss, K., Kollmer, I., Todt, D., 2004. Acoustic communication in noise: regulation of call characteristics in a New World monkey. Journal of Experimental Biology 207, 443–448.

Cappe, C., Barone, P., 2005. Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. European Journal of Neuroscience 22, 2886–2902.

Cappe, C., Thut, G., Romei, V., Murray, M.M., 2009. Selective integration of auditory-visual looming cues by humans. Neuropsycholgia 47, 1045–1052.

Chandrasekaran, C., Ghazanfar, A.A., 2009. Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. Journal of Neurophysiology 101, 773–788.

Cheney, D.L., Seyfarth, R.M., 1982. How Vervet monkeys perceive their grunts – field playback experiments. Animal Behaviour 30, 739–751.

Clark, A., 1997. Being There: Putting Brain, Body, and World Together Again. MIT Press, Cambridge, MA.

de la Mothe, L.A., Blumell, S., Kajikawa, Y., Hackett, T.A., 2006. Cortical connections of the auditory cortex in marmoset monkeys: Core and medial belt regions. Journal of Comparative Neurology 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.

Egnor, S.E.R., Hauser, M.D., 2006. Noise-induced vocal modulation in cotton-top tamarins (Saguinus oedipus). American Journal of Primatology 68, 1183–1190.

Egnor, S.E.R., Iguina, C.G., Hauser, M.D., 2006. Perturbation of auditory feedback causes systematic perturbation in vocal structure in adult cotton-top tamarins. Journal of Experimental Biology 209, 3652–3663.

Egnor, S.E.R., Wickelgren, J.G., Hauser, M.D., 2007. Tracking silence: adjusting vocal production to avoid acoustic interference. Journal of Comparative Physiology A – Neuroethology Sensory Neural and Behavioral Physiology 193, 477–483.

Eliades, S.J., Wang, X.Q., 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453, 1102–1107.

Ettlinger, G., Wilson, W.A., 1990. Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms. Behavioural Brain Research 40, 169–192.

Evans, T.A., Howell, S., Westergaard, G.C., 2005. Auditory-visual cross-modal perception of communicative stimuli in tufted capuchin monkeys (Cebus apella). Journal of Experimental Psychology – Animal Behavior Processes 31, 399–406.

Fitch, W.T., 1997. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. Journal of the Acoustical Society of America 102, 1213–1222.

Fitch, W.T., Hauser, M.D., 1995. Vocal production in nonhuman-primates – acoustics, physiology, and functional constraints on honest advertisement. American Journal of Primatology 37, 191–219.

Fowler, C.A., 2004. Speech as a supramodal or amodal phenomenon. In: Calvert, G.A., Spence, C., Stein, B.E. (Eds.), The Handbook of Multisensory Processes. MIT Press, Cambridge, MA., pp. 189–201.

Fu, K.M.G., Johnston, T.A., Shah, A.S., Arnold, L., Smiley, J., Hackett, T.A., Garraghty, P.E., Schroeder, C.E., 2003. Auditory cortical neurons respond to somatosensory stimulation. Journal of Neuroscience 23, 7510–7515.

Fu, K.M.G., Shah, A.S., O'Connell, M.N., McGinnis, T., Eckholdt, H., Lakatos, P., Smiley, J., Schroeder, C.E., 2004. Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. Journal of Neurophysiology 92, 3522–3531.

Ghazanfar, A.A., Hauser, M.D., 1999. The neuroethology of primate vocal communication: substrates for the evolution of speech. Trends in Cognitive Sciences 3, 377–384.

Ghazanfar, A.A., Hauser, M.D., 2001. The auditory behaviour of primates: a neuroethological perspective. Current Opinion in Neurobiology 11, 712–720.

Ghazanfar, A.A., Logothetis, N.K., 2003. Facial expressions linked to monkey calls. Nature 423, 937–938.

Ghazanfar, A.A., Schroeder, C.E., 2006. Is neocortex essentially multisensory? Trends in Cognitive Sciences 10, 278–285.

Ghazanfar, A.A., Chandrasekaran, C.F., 2007. Paving the way forward: integrating the senses through phase-resetting of cortical oscillations. Neuron 53, 162–164.

Ghazanfar, A.A., Rendall, D., 2008. Evolution of human vocal production. Current Biology 18, R457–R460.

Ghazanfar, A.A., Nielsen, K., Logothetis, N.K., 2006. Eye movements of monkeys viewing vocalizing conspecifics. Cognition 101, 515–529.

Ghazanfar, A.A., Chandrasekaran, C., Logothetis, N.K., 2008. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. Journal of Neuroscience 28, 4457–4469.

Ghazanfar, A.A., Flombaum, J.I., Miller, C.T., Hauser, M.D., 2001. The units of perception in the antiphonal calling behavior of cotton-top tamarins (Saguinus oedipus): playback experiments with long calls. Journal of Comparative Physiology A – Neuroethology Sensory Neural and Behavioral Physiology 187, 27–35.

Ghazanfar, A.A., Smith-Rohrberg, D., Pollen, A.A., Hauser, M.D., 2002. Temporal cues in the antiphonal long-calling behaviour of cottontop tamarins. Animal Behaviour 64, 427–438.

Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., Logothetis, N.K., 2005. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. Journal of Neuroscience 25, 5004–5012.

Ghazanfar, A.A., Turesson, H.K., Maier, J.X., van Dinther, R., Patterson, R.D., Logothetis, N.K., 2007. Vocal tract resonances as indexical cues in rhesus monkeys. Current Biology 17, 425–430.

Gogate, L.J., Walker-Andrews, A.S., Bahrick, L.E., 2001. The intersensory origins of word comprehension: an ecological-dynamic systems view. Developmental Science 4, 1–18.

Gordon, M.S., Rosenblum, L.D., 2005. Effects of intrastimulus modality change on audiovisual time-to-arrival judgments. Perception & Psychophysics 67, 580–594.

Gothard, K.M., Battaglia, F.P., Erickson, C.A., Spitler, K.M., Amaral, D.G., 2007. Neural responses to facial expression and face identity in the monkey amygdala. Journal of Neurophysiology 97, 1671–1683.

Graziano, M.S., Aflalo, T.N., 2007. Rethinking cortical organization: moving away from discrete areas arranged in hierarchies. The Neuroscientist 13, 138–147.

Hackett, T.A., Stepniewska, I., Kaas, J.H., 1998. Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. Journal of Comparative Neurology 394, 475–495.

Hackett, T.A., Stepniewska, I., Kaas, J.H., 1999. Prefrontal connections of the parabelt auditory cortex in macaque monkeys. Brain Research 817, 45–58.

Hackett, T.A., Preuss, T.M., Kaas, J.H., 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. Journal of Comparative Neurology 441, 197–222.

Hackett, T.A., de la Mothe, L.A., Ulbert, I., Karmos, G., Smiley, J., Schroeder, C.E., 2007a. Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. Journal of Comparative Neurology 502, 924–952.

Hackett, T.A., Smiley, J.F., Ulbert, I., Karmos, G., Lakatos, P., de la Mothe, L.A., Schroeder, C.E., 2007b. Sources of somatosensory input to the caudal belt areas of auditory cortex. Perception 36, 1419–1430.

Harries, M.H., Perrett, D.I., 1991. Visual processing of faces in temporal cortex – physiological evidence for a modular organization and possible anatomical correlates. Journal of Cognitive Neuroscience 3, 9–24.

Hauser, M.D., Ybarra, M.S., 1994. The role of lip configuration in monkey vocalizations – experiments using xylocaine as a nerve block. Brain and Language 46, 232–244.

Hauser, M.D., Evans, C.S., Marler, P., 1993. The role of articulation in the production of rhesus-monkey, Macaca mulatta, vocalizations. Animal Behaviour 45, 423–433.

Ito, T., Tiede, M., Ostry, D.J., 2009. Somatosensory function in speech perception. Proceedings of the National Academy of Sciences of the United States of America 106, 1245–1248.

Iyengar, S., Qi, H., Jain, N., Kaas, J.H., 2007. Cortical and thalamic connections of the representations of the teeth and tongue in somatosensory cortex of new world monkeys. Journal of Comparative Neurology 501, 95–120.

Izumi, A., Kojima, S., 2004. Matching vocalizations to vocalizing faces in a chimpanzee (Pan troglodytes). Animal Cognition 7, 179–184.

Jiang, J.T., Alwan, A., Keating, P.A., Auer, E.T., Bernstein, L.E., 2002. On the relationship between face movements, tongue movements, and speech acoustics. Eurasip Journal on Applied Signal Processing 2002, 1174–1188.

Jones, J.A., Munhall, K.G., 2003. Learning to produce speech with an altered vocal tract: the role of auditory feedback. Journal of the Acoustical Society of America 113, 532–543.

Jones, J.A., Munhall, K.G., 2005. Remapping auditory-motor representations in voice production. Current Biology 15, 1768–1772.

Jordan, K.E., Brannon, E.M., Logothetis, N.K., Ghazanfar, A.A., 2005. Monkeys match the number of voices they hear with the number of faces they see. Current Biology 15, 1034–1038.

Kayser, C., Logothetis, N.K. in press. Directed interactions between auditory and superior temporal cortices and their role in sensory integration. Frontiers in Integrative Neuroscience.

Kayser, C., Petkov, C.I., Logothetis, N.K., 2008. Visual modulation of neurons in auditory cortex. Cerebral Cortex 18, 1560–1574.

Kayser, C., Petkov, C.I., Augath, M., Logothetis, N.K., 2005. Integration of touch and sound in auditory cortex. Neuron 48, 373–384.

Kayser, C., Petkov, C.I., Augath, M., Logothetis, N.K., 2007. Functional imaging reveals visual modulation of specific fields in auditory cortex. Journal of Neuroscience 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. Archives of General Psychiatry 59, 809–816.

Kuhl, P.K., Williams, K.A., Meltzoff, A.N., 1991. Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. Journal of Experimental Psychology: Human Perception and Performance. 17, 829–840.

Kuraoka, K., Nakamura, K., 2007. Responses of single neurons in monkey amygdala to facial and vocal emotions. Journal of Neurophysiology 97, 1379–1387.

Lakatos, P., Chen, C.-M., O'Connell, M.N., Mills, A., Schroeder, C.E., 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53, 279–292.

Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., Schroeder, C.E., 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. Journal of Neurophysiology 94, 1904–1911.

Lansing, I.R., McConkie, G.W., 2003. Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. Perception & Psychophysics 65, 536–552.

Liberman, A.M., Mattingly, I., 1985. The motor theory revised. Cognition 21, 1–36.

Maier, J.X., Chandrasekaran, C., Ghazanfar, A.A., 2008. Integration of bimodal looming signals through neuronal coherence in the temporal lobe. Current Biology 18, 963–968.

Maier, J.X., Neuhoff, J.G., Logothetis, N.K., Ghazanfar, A.A., 2004. Multisensory integration of looming signals by rhesus monkeys. Neuron 43, 177–181.

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 229–239.

Meltzoff, A.N., Moore, M., 1997. Explaining facial imitation: a theoretical model. Early Development & Parenting 6, 179–192.

Nasir, S.M., Ostry, D.J., 2008. Speech motor learning in profoundly deaf adults. Nature Neuroscience 11, 1217–1222.

Noesselt, T., Rieger, J.W., Schoenfeld, M.A., Kanowski, M., Hinrichs, H., Heinze, H.-J., Driver, J., 2007. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. Journal of Neuroscience 27, 11431–11441.

Oram, M.W., Perrett, D.I., 1994. Responses of anterior superior temporal polysensory (STPa) neurons to biological motion stimuli. Journal of Cognitive Neuroscience 6, 99–116.

Palombit, R.A., Cheney, D.L., Seyfarth, R.M., 1999. Male grunts as mediators of social interaction with females in wild chacma baboons (*Papio cynocephalus ursinus*). Behaviour 136, 221–242.

Parr, L.A., 2004. Perceptual biases for multimodal cues in chimpanzee (*Pan troglodytes*) affect recognition. Animal Cognition 7, 171–178.

Patterson, M.L., Werker, J.F., 2003. Two-month-old infants match phonetic information in lips and voice. Developmental Science 6, 191–196.

Petkov, C.I., Kayser, C., Augath, M., Logothetis, N.K., 2006. Functional imaging reveals numerous fields in the monkey auditory cortex. PLoS Biology 4, 1213–1226.

Rauschecker, J.P., Tian, B., Hauser, M., 1995. Processing of complex sounds in the macaque nonprimary auditory-cortex. Science 268, 111–114.

Recanzone, G.H., 2008. Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. Journal of Neuroscience 28, 13184–13193.

Remedios, R., Logothetis, N.K., Kayser, C., 2009. An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. Journal of Neuroscience 29, 1034–1045.

Robinson, D.A., Fuchs, A.F., 1969. Eye movements evoked by stimulation of frontal eye fields. Journal of Neurophysiology 32, 637–648.

Romanski, L.M., Ghazanfar, A.A. in press. The primate frontal and temporal lobes and their role in multisensory vocal communication. In: Platt, M.L., Ghazanfar, A.A. (Eds.), Primate Neuroethology. Oxford University Press, Oxford.

Romanski, L.M., Bates, J.F., Goldman-Rakic, P.S., 1999. Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. Journal of Comparative Neurology 403, 141–157.

Romanski, L.M., Averbeck, B.B., Diltz, M., 2005. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. Journal of Neurophysiology 93, 734–747.

Rosenblum, L.D., 2005. Primacy of multimodal speech perception. In: Pisoni, D.B., Remez, R.E. (Eds.), Handbook of Speech Perception. Blackwell, Malden, MA, pp. 51–78.

Sams, M., Mottonen, R., Sihvonen, T., 2005. Seeing and hearing others and oneself talk. Cognitive Brain Research 23, 429–435.

Schall, J.D., Morel, A., King, D.J., Bullier, J., 1995. Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. Journal of Neuroscience 15, 4464–4487.

Schroeder, C.E., Foxe, J.J., 2002. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cognitive Brain Research 14, 187–198.

Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., Puce, A., 2008. Neuronal oscillations and visual amplification of speech. Trends in Cognitive Sciences 12, 106–113.

Schroeder, C.E., Lindsley, R.W., Specht, C., Marcovici, A., Smiley, J.F., Javitt, D.C., 2001. Somatosensory input to auditory association cortex in the macaque monkey. Journal of Neurophysiology 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, D.N., 1989. Frontal-lobe connections of the superior temporal sulcus in the rhesus-monkey. Journal of Comparative Neurology 281, 97–113.

Seltzer, B., Pandya, D.N., 1994. Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. Journal of Comparative Neurology 343, 445–463.

Sinnott, J.M., Stebbins, W.C., Moody, D.B., 1975. Regulation of voice amplitude by monkey. Journal of the Acoustical Society of America 58, 412–414.

Smiley, J.F., Hackett, T.A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D.C., Schroeder, C.E., 2007. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. Journal of Comparative Neurology 502, 894–923.

Sugihara, T., Diltz, M.D., Averbeck, B.B., Romanski, L.M., 2006. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. Journal of Neuroscience 26, 11138–11147.

Tremblay, S., Shiller, D.M., Ostry, D.J., 2003. Somatosensory basis of speech production. Nature 423, 866–869.

van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. Proceedings of the National Academy of Sciences of the United States of America 102, 1181–1186.

Werner-Reiss, U., Kelly, K.A., Trause, A.S., Underhill, A.M., Groh, J.M., 2003. Eye position affects activity in primary auditory cortex of primates. Current Biology 13, 554–562.

Yehia, H., Rubin, P., Vatikiotis-Bateson, E., 1998. Quantitative association of vocal-tract and facial behavior. Speech Communication 26, 23–43.

Yehia, H.C., Kuratate, T., Vatikiotis-Bateson, E., 2002. Linking facial animation, head motion and speech acoustics. Journal of Phonetics 30, 555–568.