



HHS Public Access

Author manuscript

J Cogn Neurosci. Author manuscript; available in PMC 2015 May 22.

Published in final edited form as:

J Cogn Neurosci. 2014 June ; 26(6): 1196–1207. doi:10.1162/jocn_a_00575.

Facial Expressions and the Evolution of the Speech Rhythm

Asif A. Ghazanfar and Daniel Y. Takahashi

Princeton University

Abstract

In primates, different vocalizations are produced, at least in part, by making different facial expressions. Not surprisingly, humans, apes, and monkeys all recognize the correspondence between vocalizations and the facial postures associated with them. However, one major dissimilarity between monkey vocalizations and human speech is that, in the latter, the acoustic output and associated movements of the mouth are both rhythmic (in the 3- to 8-Hz range) and tightly correlated, whereas monkey vocalizations have a similar acoustic rhythmicity but lack the concomitant rhythmic facial motion. This raises the question of how we evolved from a presumptive ancestral acoustic-only vocal rhythm to the one that is audiovisual with improved perceptual sensitivity. According to one hypothesis, this bisensory speech rhythm evolved through the rhythmic facial expressions of ancestral primates. If this hypothesis has any validity, we expect that the extant nonhuman primates produce at least some facial expressions with a speech-like rhythm in the 3- to 8-Hz frequency range. Lip smacking, an affiliative signal observed in many genera of primates, satisfies this criterion. We review a series of studies using developmental, x-ray cineradiographic, EMG, and perceptual approaches with macaque monkeys producing lip smacks to further investigate this hypothesis. We then explore its putative neural basis and remark on important differences between lip smacking and speech production. Overall, the data support the hypothesis that lip smacking may have been an ancestral expression that was linked to vocal output to produce the original rhythmic audiovisual speech-like utterances in the human lineage.

INTRODUCTION

Both speech and nonhuman primate vocalizations are produced by the coordinated movements of the lungs, larynx (vocal folds), and the supralaryngeal vocal tract (Ghazanfar & Rendall, 2008; Fitch & Hauser, 1995). The vocal tract consists of the pharynx, mouth, and nasal cavity through which a column of air is produced. The shape of this column determines its resonance properties and thus in which frequency bands of the sound produced at the laryngeal source get emphasized or suppressed. During vocal production, the shape of the vocal tract can be changed by moving the various effectors of the face (including the lips, jaw, and tongue) into different positions. The different shapes, along with changes in vocal fold tension and respiratory power, are what give rise to different sounding vocalizations. To put it simply: Different vocalizations (including different speech sounds) are produced in part by making different facial expressions.

© 2014 Massachusetts Institute of Technology

Reprint requests should be sent to Asif A. Ghazanfar, Neuroscience Institute, Department of Psychology, Princeton University, Princeton, NJ 08540, or via asifg@princeton.edu.

Vocal tract motion not only changes the acoustics of vocalizations by changing their resonance frequencies but also results in the predictable deformation of the face around the mouth and other parts of the face (Yehia, Kuratate, & Vatikiotis-Bateson, 2002; Yehia, Rubin, & Vatikiotis-Bateson, 1998; Hauser & Ybarra, 1994; Hauser, Evans, & Marler, 1993). Different macaque monkey (*Macaca* spp.) vocalizations are produced with unique lip configurations and mandibular positions, and the motion of such articulators influences the acoustics of the signal (Hauser & Ybarra, 1994; Hauser et al., 1993). For example, coo calls, like /u/ in speech, are produced with the lips protruded, whereas screams, like the /i/ in speech, are produced with the lips retracted (Figure 1). Facial motion cues used by humans for speech reading are present during primate vocal production as well. The fact that different vocalizations are produced through different facial expressions and are therefore inherently “multisensory” is typically ignored by theories regarding the evolution of speech/language that focus solely on laryngeal control by the neocortex (or the lack thereof; Arbib, 2005; Jarvis, 2004).

Naturally, any vertebrate organism (from fishes and frogs to birds and dogs) that produces vocalizations will have a simple, concomitant visible motion in the area of the mouth. However, in the primate lineage, both the number and diversity of muscles innervating the face (Burrows, Waller, & Parr, 2009; Huber, 1930a, 1930b) and the amount of neural control related to facial movement (Sherwood, 2005; Sherwood et al., 2005; Sherwood, Holloway, Erwin, & Hof, 2004; Sherwood, Holloway, Erwin, Schleicher, et al., 2004) increased over the course of evolution relative to other taxa. This increase in the number of muscles allowed the production of a greater diversity of facial and vocal expressions in primates (Andrew, 1962). The inextricable link between vocal output and facial expressions allows many nonhuman primates to 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, Howell, & Westergaard, 2005; Izumi & Kojima, 2004; Parr, 2004; Ghazanfar & Logothetis, 2003). For example, macaque monkeys, without training or reward, match individual identity and expression types across modalities (Sliwa, Duhamel, Pascalis, & Wirth, 2011; Ghazanfar & Logothetis, 2003), segregate competing voices in noisy conditions using vision (Jordan, Brannon, Logothetis, & Ghazanfar, 2005), and use formant frequencies to estimate the body size of conspecifics (Ghazanfar et al., 2007). More recently, monkeys trained to detect vocalizations in noise demonstrated that seeing concomitant facial motion sped up their RTs in a manner identical to that of humans detecting speech sounds (Chandrasekaran, Lemus, Trubanova, Gondan, & Ghazanfar, 2011).

There are also some very important differences in how humans versus nonhuman primates produce their utterances (Ghazanfar & Rendall, 2008), and these differences further enhance human multisensory communication above and beyond what monkeys can do. One universal feature of speech—typically lacking in monkey vocalizations—is its bisensory rhythm. That is, when humans speak, both the acoustic output and the movements of the mouth are highly rhythmic and tightly correlated with each other (Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009).

THE HUMAN SPEECH RHYTHM VERSUS MONKEY VOCALIZATIONS

Across all languages studied to date, speech typically exhibits a 3- to 8-Hz rhythm that is, for the most part, related to the rate of syllable production (Chandrasekaran et al., 2009; Greenberg, Carvey, Hitchcock, & Chang, 2003; Crystal & House, 1982; Malecot, Johnson, & Kizziar, 1972; Figure 2A). Both mouth motion and the acoustic envelope of speech are rhythmic. This 3- to 8-Hz rhythm is critical to speech perception. Disrupting the acoustic component of this rhythm significantly reduces intelligibility (Elliot & Theunissen, 2009; Ghitza & Greenberg, 2009; Smith, Delgutte, & Oxenham, 2002; Saberi & Perrott, 1999; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Drullman, Festen, & Plomp, 1994), as does disrupting the visual components arising from facial movements (Vitkovitch & Barber, 1996). Thus, the speech rhythm parses the signal into basic units from which information on a finer (faster) temporal scale can be extracted (Ghitza, 2011). Given the importance of this rhythm in speech and its underlying neurophysiology (Ghazanfar & Poeppel, in press; Giraud & Poeppel, 2012), understanding how speech evolved requires investigating the origins of its bisensory rhythmic structure.

Oddly enough, macaque monkey vocalizations have a similar acoustic rhythmicity but without the concomitant and temporally correlated rhythmic facial motion (Figure 2B). Macaque vocalizations are typically produced with a single ballistic facial motion. Modulation spectra analyses of the acoustic rhythmicity of macaque monkey vocalizations longer than 400 msec reveals that their rhythmicity is strikingly similar to that of the acoustic envelope for speech (Figure 3A). Both signals fall within the 3- to 8-Hz range (see also Cohen, Theunissen, Russ, & Gill, 2007, for shared low-frequency components of macaque monkey calls and speech). Moreover, examination of a single call category (Figure 3B, top) or an exemplar (Figure 3B, bottom) shows that this rhythmicity is not the result of averaging across call-type categories or averaging within a single call category. Thus, one key evolutionary question is: How did we evolve from a presumptive ancestral unisensory, acoustic-only vocal rhythm (Figure 2B) to the one that is audiovisual, with both mouth movements and acoustics sharing the same rhythmicity (Figure 2A)?

ON THE ORIGINS OF THE SPEECH RHYTHM

One theory posits that the rhythm of speech evolved through the modification of rhythmic facial movements in ancestral primates (MacNeilage, 1998, 2008). In extant primates, such facial movements are extremely common as visual communicative gestures. Lip smacking, for example, is an affiliative signal observed in many genera of primates (Redican, 1975; Hinde & Rowell, 1962; Van Hooff, 1962), including chimpanzees (Parr, Cohen, & de Waal, 2005). It is characterized by regular cycles of vertical jaw movement, often involving a parting of the lips but sometimes occurring with closed, puckered lips. Although lip smacking by both monkeys and chimpanzees is often produced during grooming interactions, monkeys also exchange lip-smacking bouts during face-to-face interactions (Ferrari, Paukner, Ionica, & Suomi, 2009; Van Hooff, 1962). Moreover, lip-smacking gestures are among the first facial expressions produced by infant monkeys (De Marco & Visalberghi, 2007; Ferrari et al., 2006) and are frequently used during mother–infant interactions (Ferrari et al., 2009). According to MacNeilage (1998, 2008), during the course

of speech evolution, such nonvocal rhythmic facial expressions were coupled to vocalizations to produce the audiovisual components of babbling-like (i.e., consonant–vowel-like) speech expressions in the human lineage (Figure 2C).

Although direct tests of such evolutionary hypotheses are difficult, we can use the 3- to 8-Hz rhythmic signature of speech as a foundation to explore its veracity. There are now many lines of evidence that demonstrate that the production of lip smacking in macaque monkeys is similar to the orofacial rhythms produced during speech. First and foremost, lip smacking exhibits a speech-like rhythm in the 3- to 8-Hz frequency range (Ghazanfar, Chandrasekaran, & Morrill, 2010). This rhythmic frequency range is distinct from that of chewing and teeth grinding (an anxiety-driven expression), although all three rhythmic orofacial motions use the same effectors. Yet, it still may be that the 3- to 8-Hz range is large enough that the correspondence between the speech rhythm and the lip-smacking rhythm is coincidental. Below, we provide evidence from development, x-ray cineradiography, EMG, and perception that suggests otherwise.

Developmental Parallels

If the underlying mechanisms that produce the rhythm in monkey lip smacking and human speech are homologous, then their developmental trajectories should be similar (Gottlieb, 1992; Schneirla, 1949). Moreover, this common trajectory should be distinct from the developmental trajectory of other rhythmic mouth movements. In humans, the earliest form of rhythmic and voluntary vocal behavior occurs some time after 6 months of age, when vocal babbling abruptly emerges (Preuschoff, Quartz, & Bossaerts, 2008; Smith & Zelaznik, 2004; Locke, 1993). Babbling is characterized by the production of canonical syllables that have acoustic characteristics similar to adult speech. Their production involves rhythmic sequences of a mouth close–open alternation (Oller, 2000; Lindblom, Krull, & Stark, 1996; Davis & MacNeilage, 1995). This close–open alternation results in a consonant–vowel syllable representing the only syllable type present in all the world’s languages (Bell & Hooper, 1978). However, babbling does not emerge with the same rhythmic structure as adult speech, but rather, there is a sequence of structural changes in the rhythm. There are two main aspects to these changes: frequency and variability. In adults, the speech rhythm is ~5 Hz (Chandrasekaran et al., 2009; Dolata, Davis, & MacNeilage, 2008; Greenberg et al., 2003; Crystal & House, 1982; Malecot et al., 1972), whereas in infant babbling, the rhythm is considerably slower. Infants produce speech-like sounds at a slower rate of roughly 2.8–3.4 Hz (Dolata et al., 2008; Nathani, Oller, & Cobo-Lewis, 2003; Lynch, Oller, Steffens, & Buder, 1995; Levitt & Wang, 1991). In addition to differences in the rhythmic frequency between adults and infants, there are differences in their variability. Infants produce highly variable vocal rhythms (Dolata et al., 2008) that do not become fully adult-like until postpubescence (Smith & Zelaznik, 2004). Importantly, this developmental trajectory from babbling to speech is distinct from that of another cyclical mouth movement—chewing. The frequency of chewing movements in humans is highly stereotyped: It is slow in frequency and remains virtually unchanged from early infancy into adulthood (Green et al., 1997; Kiliaridis, Karlsson, & Kjellberge, 1991). Chewing movements are often used as reference movement in speech production studies because both movements use the very same effectors.

By measuring the rhythmic frequency and variability of lip smacking in macaque monkeys across neonatal, juvenile, and adult age groups, the hypothesis that lip smacking develops in the same way as the speech rhythm was tested (Morrill, Paukner, Ferrari, & Ghazanfar, 2012). There were at least three possible outcomes. First, given the differences in the size of the orofacial structures between macaques and humans (Ross et al., 2009), it is possible that lip-smacking and speech rhythms do not converge on the same ~5-Hz rhythm. Second, because of the precocious neocortical development of macaque monkeys relative to humans (Malkova, Heuer, & Saunders, 2006; Gibson, 1991), the lip-smacking rhythm could remain stable from birth onwards and show no changes in frequency and/or variability (much like the chewing rhythm in humans; Thelen, 1981). Finally, lip-smacking dynamics may undergo the same developmental trajectory as the human speech rhythm: decreasing variability, with increasing frequency converging onto a ~5-Hz rhythm.

The developmental trajectory of monkey lip smacking parallels speech development (Morrill et al., 2012; Locke, 2008). Measurements of the rhythmic frequency and variability of lip smacking across individuals in three different age groups (neonates, juveniles, and adults) revealed that young individuals produce slower, more variable mouth movements, and as they get older, these movements become faster and less variable (Morrill et al., 2012)—this is exactly as speech develops, from babbling to adult consonant–vowel production (Dolata et al., 2008). The developmental trajectory for lip smacking was distinct from that of chewing (Morrill et al., 2012). As in humans (Green et al., 1997; Kiliaridis et al., 1991), chewing had the same slow frequency and consistent low variability across age groups. Thus, these differences in developmental trajectories between lip smacking and chewing are identical to those reported in humans for speech and chewing (Steeve, 2010; Steeve, Moore, Green, Reilly, & McMurtrey, 2008; Moore & Ruark, 1996).

The Coordination of Effectors

If the human speech and monkey lip smacking have a shared neural basis, one would expect commonalities in the coordination of the effectors involved. One piece of evidence for this comes from motor control. During speech, different sounds are produced through the coordination between key vocal tract anatomical structures: the jaw/lips, tongue, and hyoid. [The hyoid is a bony structure to which the laryngeal muscles attach.] These effectors are more loosely coupled during speech movements than during chewing movements (Matsuo & Palmer, 2010; Hiimae & Palmer, 2003; Hiimae et al., 2002; Ostry & Munhall, 1994; Moore, Smith, & Ringel, 1988). X-ray cineradiography (x-ray movies) used to visualize the internal dynamics of the macaque monkey vocal tract during lip smacking and chewing revealed that lips, tongue, and hyoid move during lip smacks (as in speech) and do so with a speech-like 3- to 8-Hz rhythm (Figure 4A and B). Relative to lip smacking, movements during chewing were significantly slower for each of these structures (Figure 4B). The temporal coordination of these structures was distinct for each behavior (Figure 4C). Partial directed coherence measures—an analysis that measures to what extent one time series can predict another (Takahashi, Baccala, & Sameshima, 2010)—revealed that, although the hyoid moves continuously during lip smacking, there is no coupling of the hyoid with lip and tongue movements; whereas during chewing, we observed more coordination between the three structures. These patterns are consistent with what is observed during human

speech and chewing (Hiemae & Palmer, 2003; Hiemae et al., 2002): The effectors are more loosely coupled during lip smacking than during chewing. Furthermore, the spatial displacement of the lips, tongue, and hyoid is greater during chewing than during lip smacking (Ghazanfar, Takahashi, Mathur, & Fitch, 2012), again similar to what is observed in human speech versus chewing (Hiemae et al., 2002).

Facial EMG studies of muscle coordination during lip smacking and chewing also revealed very distinct activity patterns associated with each behavior (Shepherd, Lanzilotto, & Ghazanfar, 2012). The coordination of monkeys' orofacial musculature during lip smacking and chewing was measured using EMG electrodes targeting five muscles—three in the lower face (zygomaticus, orbicularis oris, and mentalis) and two in the upper face (frontalis and auricularis; Figure 5A and B). Muscle coordination was evident in both lip-smacking and chewing behavior, but the coordination of the perioral muscles was stronger and more stereotyped during lip smacking than during chewing. Whereas lip smacking is characterized by coherent movements of perioral mimetic muscles, chewing exhibits inconsistent perioral coordination in these muscles despite strong coordination of signal at the auricular and zygomatic sites (Figure 5C). These data suggest that lip smacking (like speech) has a distinct motor program. It is not simply a ritualization of feeding behavior although it may have evolved from such a behavior following a reorganization in the motor program. Reorganization of central pattern generators underlying rhythmic behaviors is not uncommon in nervous system evolution (Newcomb, Sakurai, Lillvis, Gunaratne, & Katz, 2012).

Perceptual Tuning

In speech, disrupting the auditory or visual component of the 3- to 8-Hz rhythm significantly reduces intelligibility (Elliot & Theunissen, 2009; Smith et al., 2002; Saberi & Perrott, 1999; Vitkovitch & Barber, 1996; Shannon et al., 1995; Drullman et al., 1994). To test whether monkeys were differentially sensitive to lip smacking produced with a rhythmic frequency in the species-typical range (mean = 4–6 Hz; Ghazanfar et al., 2010, 2012; Morrill et al., 2012), a preferential-looking procedure was used (Ghazanfar, Morrill, & Kayser, 2013). Computer-generated monkey avatars were used to produce stimuli varying in lip-smacking frequency within (6 Hz) and outside (3 and 10 Hz) the species-typical range but with otherwise identical features (Chandrasekaran et al., 2011; Steckenfinger & Ghazanfar, 2009; Figure 6A). The use of avatar faces allowed control of additional factors that could potentially influence attention, such as head and eye movements and lighting conditions for face and background. Each of two unique avatar faces was generated to produce the three different lip-smacking rhythms.

Measuring looking times to one or the other avatar assessed preferential looking. There were at least five possible outcomes. First, monkeys could show no preference at all, suggesting that they either did not find the avatars salient, that they failed to discriminate the different frequencies, or that they preferred one of the avatar identities (as opposed to the lip-smacking rhythm) over the others. Second, they could show a preference for slower lip-smacking rhythms (3 Hz > 6 Hz > 10 Hz). Third, they could prefer faster rhythms (3 Hz < 6 Hz < 10 Hz; Lewkowicz, 1985). Fourth, they could show avoidance of the 6-Hz lip

smacking, preferring the unnatural 3- and 10-Hz rhythms to the natural lip-smacking rhythm. This may arise if monkeys find the naturalistic 6-Hz lip smacking disturbing (perhaps uncanny; Steckenfinger & Ghazanfar, 2009) or too arousing (Zangenehpour, Ghazanfar, Lewkowicz, & Zatorre, 2009). Finally, monkeys could show a preference for the 6-Hz lip smacking over the 3 and 10 Hz, perhaps because such a rhythm is concordant with the rhythmic activity patterns in the neocortex (Giraud & Poeppel, 2012; Karmel, Lester, McCarvill, Brown, & Hofmann, 1977). Monkeys showed an overall preference for the natural 6-Hz rhythm when compared with the perturbed rhythms (Figure 6B). This lends behavioral support for the hypothesis that their perceptual processes are similarly tuned to the natural frequencies of communication signals as they are for the speech rhythm in humans.

NEURAL MECHANISMS

These multisensory, developmental, bio-mechanical, and perceptual homologies between human speech and monkey lip smacking suggest that their underlying neural mechanisms for perception and production of communication signals may also be homologous. On the sensory-perception side, neurophysiological work in the inferior temporal lobe demonstrated that neurons in that area are “face sensitive” (Gross, Rocha-Miranda, & Bender, 1972; Gross, Bender, & Rocha-Miranda, 1969) and multisensory (Bruce, Desimone, & Gross, 1981), and it is areas within this region, along with the auditory cortex, that are activated by audiovisual speech in humans (Arnal, Morillon, Kell, & Giraud, 2009; von Kriegstein et al., 2008; van Wassenhove, Grant, & Poeppel, 2005; Callan et al., 2003; Calvert et al., 1999). Similarly, in macaque monkeys, neurons in the STS, pFC, and auditory cortex are driven and modulated by species-specific audiovisual communication signals (Ghazanfar et al., 2010; Chandrasekaran & Ghazanfar, 2009; Ghazanfar, Chandrasekaran, & Logothetis, 2008; Sugihara, Diltz, Averbek, & Romanski, 2006; Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Ghazanfar, Maier, Hoffman, & Logothetis, 2005). Neurons in the STS are also very sensitive to dynamic faces, including lip-smacking gestures (Ghazanfar et al., 2010). Much of this comparative multisensory work has been reviewed elsewhere (Ghazanfar & Chandrasekaran, 2012) as has the relationship between the speech rhythm, speech perception, and auditory cortical oscillations (Giraud & Poeppel, 2012).

Much less is known about the neural mechanisms underlying the production of rhythmic communication signals in human and nonhuman primates. The mandibular movements shared by chewing, lip smacking, vocalizations, and speech all require the coordination of muscles controlling the jaw, face, tongue, and respiration. Their foundational rhythms are likely produced by homologous central pattern generators in the pons and medulla of the brainstem (Lund & Kolta, 2006). These circuits are present in all mammals, are operational early in life, and are modulated by feedback from peripheral sensory receptors. Beyond peripheral sensory feedback, the neocortex is an additional source influencing how differences (e.g., frequency and variability) between orofacial movements may arise (Lund & Kolta, 2006; MacNeilage, 1998). Whereas chewing movements may be largely independent of cortical control (Lund & Kolta, 2006), lip smacking and speech production are both modulated by the neocortex, in accord with social context and communication goals (Caruana, Jezzini, Sbriscia-Fioretti, Rizzolatti, & Gallese, 2011; Bohland & Guenther,

2006). Thus, one hypothesis for the developmental changes in the frequency and variability of lip smacking and speech is that they are a reflection of the development of neocortical circuits influencing brainstem central pattern generators.

One important neocortical node likely to be involved in this circuit is the insula. The human insula is involved in multiple processes related to communication, including feelings of empathy (Keysers & Gazzola, 2006) and learning in highly dynamic social environments (Preuschoff et al., 2008). Importantly, the human insula is also involved in speech production (Bohland & Guenther, 2006; Ackermann & Riecker, 2004; Catrin Blank, Scott, Murphy, Warburton, & Wise, 2002; Dronkers, 1996). Consistent with an evolutionary link between lip smacking and speech, the insula also plays a role in generating monkey lip smacking (Caruana et al., 2011). Electrical stimulation of the insula elicits lip smacking in monkeys, but only when those monkeys are making eye contact (i.e., are face-to-face) with another individual. This demonstrates that the insula is a social sensory-motor node for lip-smacking production. Thus, it is conceivable that, for both monkey lip smacking and human speech, the increase in rhythmic frequency and decrease in variability are, in part at least, because of the socially guided development of the insula. Another possible cortical node in this network is the premotor cortex in which neurons respond to seeing and producing lip-smacking expressions (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003).

A neural mechanism is needed to link lip-smacking-like facial expressions to concomitant vocal output (the laryngeal source). This separate origin of laryngeal control remains a mystery. A plausible scenario is that the cortical control of the brainstem's nucleus ambiguus, which innervates the laryngeal muscles, is absent in all primates save humans (Deacon, 1997).

DIFFERENCES BETWEEN LIP SMACKING AND SPEECH PRODUCTION

Two core features of speech production—its rhythmical structure and temporal coordination of vocal tract effectors—are shared with lip smacking. Yet, there are striking differences between the two modes of expression, the most obvious of which is that lip smacking lacks a vocal (laryngeal) component. Although a quiet consonant-like bilabial plosive or /p/ sound is produced when the lips smack together, no sound is generated by the larynx. Thus, the capacity to produce vocalizations during rhythmic vocal tract movements seen in speech seems to be a human adaptation. How can lip smacking be related to speech if there is no vocal component? In human and nonhuman primates, the basic mechanisms of voice production are broadly similar and consist of two distinct components: the laryngeal source and the vocal tract filter (Ghazanfar & Rendall, 2008; Fitch & Hauser, 1995; Fant, 1970). Voice production involves (1) a sound generated by air pushed by the lungs through the vibrating vocal folds within the larynx (the source) and (2) the modification through linear filtering of this sound by the vocal tract airways above the larynx (the filter). The filter consists of the nasal and oral cavities whose shapes can be changed by movements of the jaw, tongue, hyoid, and lips. These two basic components of the vocal apparatus behave and interact in complex ways to generate a wide range of sounds. The lip-smacking hypothesis for the evolution of rhythmic speech and the data that support it only address the evolution of vocal tract movements (the filter component) involved in speech production.

Other differences between lip smacking and speech are the coupling of the lips with the tongue and the range of hyoid movements (Ghazanfar et al., 2012). The coupling of the lips and tongue during lip smacking (Figure 4C) is unlikely to occur during human speech where the independence of these effectors allows for the production of a wide range of sounds (although this has not been tested explicitly). With regard to the range of hyoid movements, the hyoid occupies the same active space during lip smacking and chewing, whereas cine-radiography studies of human speech versus chewing show a dichotomy in hyoid movement patterns (Hiemae et al., 2002). These movement range differences of the hyoid in humans versus macaques could be because of functional differences in suprahyoid muscle length, the degree of neural control over this muscle group, and/or by species differences in hyoid position. During human development, the position of the hyoid relative to the mandible and tongue changes (Lieberman, McCarthy, Hiemae, & Palmer, 2001). This change allows for an increase in the range of tongue movements and, possibly, hyoid movements, relative to what is observed in nonhuman primates. Movements of either or both effectors could influence the active space of the hyoid thereby increasing the range of possible vocal tract shapes.

BRIDGING THE GAP

How easy would it be to link vocalizations to a rhythmic facial expression during the course of evolution? Recent work on gelada baboons (*Theropithecus gelada*) proves to be illuminating. Geladas are a highly specialized type of baboon. Their social structure and habitat are unique among baboons and other Old World primates as are a few of their vocalizations (Gustison, le Roux, & Bergman, 2012). One of those unique vocalizations, known as a “wobble,” is produced only by males of this species and during close, affiliative interactions with females. Wobbles are essentially lip-smacking expressions produced concurrently with vocalization (Bergman, 2013). Moreover, their rhythmicity falls within the range of the speech rhythm and lip smacking by macaque monkeys. Given that gelada baboons are very closely related to yellow baboons (their taxa are separated by 4 million years) who do not produce anything like wobble vocalizations, it appears that linking rhythmic facial expressions like lip smacking to vocal output is quite plausible. How geladas achieved this feat at the level of neural circuits is unknown, but finding out could reveal what was critical for the human transition to rhythmic audiovisual vocal output during the course of our evolution.

CONCLUSION

Human speech is not uniquely multisensory—visible facial motion is inextricably linked to acoustics. The default mode of communication in many primates is also multisensory. Apes and monkeys recognize the correspondence between vocalizations and the facial postures associated with them. One striking dissimilarity between monkey vocalizations and human speech is that the latter has a unique bisensory rhythmic structure in that both the acoustic output and the movements of the mouth are rhythmic and tightly correlated. According to one hypothesis, this bimodal speech rhythm evolved through the rhythmic facial expressions of ancestral primates. Developmental, cineradiographic, EMG, and perceptual data from macaque monkeys all support the notion that a rhythmic facial expression common among

many primate species—lip smacking—may have been one such ancestral expression. Further explorations of this hypothesis must include a broader comparative sample, especially investigations of the temporal dynamics of facial and vocal expressions in the great apes. Understanding the neural basis of both lip smacking and speech production—their similarities and differences—would also be illuminating.

Acknowledgments

A. A. G. dedicates this review to Charlie Gross. Charlie has long been a rewarding influence on me. First, as an undergraduate, when his work on the history of neuroscience captured my imagination, and later, when his pioneering studies on face-sensitive and multisensory neurons in the temporal lobe helped drive the direction of my postdoctoral research. He remains an ongoing influence on me as my close colleague, friend, and unofficial mentor at Princeton University. We thank Adrian Bartlett for the analyses shown in Figure 3 and Jeremy Borjon, Diego Cordero, and Lauren Kelly for their comments on an early version of this manuscript. This work was supported by NIH R01NS054898 (A. A. G.), the James S. McDonnell Scholar Award (A. A. G.), a Pew Latin American Fellowship (D. Y. T.), and a Brazilian Science without Borders Fellowship (D. Y. T.).

REFERENCES

- Ackermann H, Riecker A. The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language*. 2004; 89:320–328. [PubMed: 15068914]
- Andrew RJ. The origin and evolution of the calls and facial expressions of the primates. *Behaviour*. 1962; 20:1–109.
- Arbib MA. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*. 2005; 28:105–124. [PubMed: 16201457]
- Arnal LH, Morillon B, Kell CA, Giraud AL. Dual neural routing of visual facilitation in speech processing. *The Journal of Neuroscience*. 2009; 29:13445–13453. [PubMed: 19864557]
- Barracough NE, Xiao D, Baker CI, Oram MW, Perrett DI. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*. 2005; 17:377–391. [PubMed: 15813999]
- Bell, A.; Hooper, JB. *Syllables and segments*. North-Holland: Amsterdam; 1978.
- Bergman TJ. Speech-like vocalized lip-smacking in geladas. *Current Biology*. 2013; 23:R268–R269. [PubMed: 23578870]
- Bohland JW, Guenther FH. An fMRI investigation of syllable sequence production. *Neuroimage*. 2006; 2:821–841. [PubMed: 16730195]
- Bruce C, Desimone R, Gross CG. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*. 1981; 46:369–384. [PubMed: 6267219]
- Burrows AM, Waller BM, Parr LA. Facial musculature in the rhesus macaque (*Macaca mulatta*): Evolutionary and functional contexts with comparisons to chimpanzees and humans. *Journal of Anatomy*. 2009; 215:320–334. [PubMed: 19563473]
- Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E. Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport*. 2003; 14:2213–2218. [PubMed: 14625450]
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS. Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport*. 1999; 10:2619–2623. [PubMed: 10574380]
- Caruana F, Jezzini A, Sbriscia-Fioretti B, Rizzolatti G, Gallese V. Emotional and social behaviors elicited by electrical stimulation of the insula in the macaque monkey. *Current Biology*. 2011; 21:195–199. [PubMed: 21256020]
- Catrin Blank S, Scott SK, Murphy K, Warburton E, Wise RJS. Speech production: Wernicke, Broca and beyond. *Brain*. 2002; 125:1829–1838. [PubMed: 12135973]

- Chandrasekaran C, Ghazanfar AA. Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. *Journal of Neurophysiology*. 2009; 101:773–788. [PubMed: 19036867]
- Chandrasekaran C, Lemus L, Trubanova A, Gondon M, Ghazanfar AA. Monkeys and humans share a common computation for face/voice integration. *PLoS Computational Biology*. 2011; 7:e1002165. [PubMed: 21998576]
- Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA. The natural statistics of audiovisual speech. *PLoS Computational Biology*. 2009; 5:e1000436. [PubMed: 19609344]
- Cohen YE, Theunissen FE, Russ BE, Gill P. Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *Journal of Neurophysiology*. 2007; 97:1470–1484. [PubMed: 17135477]
- Crystal T, House A. Segmental durations in connected speech signals: Preliminary results. *Journal of the Acoustical Society of America*. 1982; 72:705–716. [PubMed: 7130529]
- Davis BL, MacNeilage PF. The articulatory basis of babbling. *Journal of Speech and Hearing Research*. 1995; 38:1199–1211. [PubMed: 8747814]
- De Marco A, Visalberghi E. Facial displays in young tufted capuchin monkeys (*Cebus apella*): Appearance, meaning, context and target. *Folia Primatologica*. 2007; 78:118–137.
- Deacon, TW. *The symbolic species: The coevolution of language and the brain*. New York: W.W. Norton & Company; 1997.
- Dolata JK, Davis BL, MacNeilage PF. Characteristics of the rhythmic organization of vocal babbling: Implications for an amodal linguistic rhythm. *Infant Behavior & Development*. 2008; 31:422–431. [PubMed: 18289693]
- Dronkers NF. A new brain region for coordinating speech articulation. *Nature*. 1996; 384:159–161. [PubMed: 8906789]
- Drullman R, Festen JM, Plomp R. Effect of reducing slow temporal modulations on speech reception. *Journal of the Acoustical Society of America*. 1994; 95:2670–2680. [PubMed: 8207140]
- Elliot TM, Theunissen FE. The modulation transfer function for speech intelligibility. *PLoS Computational Biology*. 2009; 5:e1000302. [PubMed: 19266016]
- Evans TA, Howell S, Westergaard GC. Auditory– visual cross-modal perception of communicative stimuli in tufted capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*. 2005; 31:399–406. [PubMed: 16248726]
- Fant, G. *Acoustic theory of speech production*. 2nd ed.. Paris: Mouton; 1970.
- Ferrari PF, Gallese V, Rizzolatti G, Fogassi L. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*. 2003; 17:1703–1714. [PubMed: 12752388]
- Ferrari PF, Paukner A, Ionica C, Suomi SJ. Reciprocal face-to-face communication between rhesus macaque mothers and their newborn infants. *Current Biology*. 2009; 19:1768–1772. [PubMed: 19818617]
- Ferrari PF, Visalberghi E, Paukner A, Fogassi L, Ruggiero A, Suomi SJ. Neonatal imitation in rhesus macaques. *PLoS Biology*. 2006; 4:1501.
- Fitch WT, Hauser MD. Vocal production in nonhuman primates—Acoustics, physiology, and functional constraints on honest advertisement. *American Journal of Primatology*. 1995; 37:191–219.
- Ghazanfar, AA.; Chandrasekaran, C. Non-human primate models of audiovisual communication. In: Stein, BE., editor. *The new handbook of multisensory processes*. Cambridge, MA: The MIT Press; 2012. p. 407-420.
- Ghazanfar AA, Chandrasekaran C, Logothetis NK. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *Journal of Neuroscience*. 2008; 28:4457–4469. [PubMed: 18434524]
- Ghazanfar AA, Chandrasekaran C, Morrill RJ. Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: Implications for the evolution of audiovisual speech. *European Journal of Neuroscience*. 2010; 31:1807–1817. [PubMed: 20584185]
- Ghazanfar AA, Logothetis NK. Facial expressions linked to monkey calls. *Nature*. 2003; 423:937–938. [PubMed: 12827188]

- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*. 2005; 25:5004–5012. [PubMed: 15901781]
- Ghazanfar AA, Morrill RJ, Kayser C. Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proceedings of the National Academy of Sciences, U.S.A.* 2013; 110:1959–1963.
- Ghazanfar, AA.; Poeppel, D. The neurophysiology and evolution of the speech rhythm. In: Gazzaniga, MS., editor. *The cognitive neurosciences V*. Cambridge, MA: The MIT Press; (in press).
- Ghazanfar AA, Rendall D. Evolution of human vocal production. *Current Biology*. 2008; 18:R457–R460. [PubMed: 18522811]
- Ghazanfar AA, Takahashi DY, Mathur N, Fitch WT. Cineradiography of monkey lipsmacking reveals the putative origins of speech dynamics. *Current Biology*. 2012; 22:1176–1182. [PubMed: 22658603]
- Ghazanfar AA, Turesson HK, Maier JX, van Dinther R, Patterson RD, Logothetis NK. Vocal tract resonances as indexical cues in rhesus monkeys. *Current Biology*. 2007; 17:425–430. [PubMed: 17320389]
- Ghitza O. Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*. 2011; 2:130. [PubMed: 21743809]
- Ghitza O, Greenberg S. On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*. 2009; 66:113–126. [PubMed: 19390234]
- Gibson, KR. Myelination and behavioral development: A comparative perspective on questions of neoteny, altriciality and intelligence. In: Gibson, KR.; Petersen, AC., editors. *Brain maturation and cognitive development: Comparative and cross-cultural perspectives*. New York: Aldine de Gruyter; 1991. p. 29–63.
- Giraud AL, Poeppel D. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*. 2012; 15:511–517.
- Gottlieb, G. *Individual development & evolution: The genesis of novel behavior*. New York: Oxford University Press; 1992.
- Green JR, Moore CA, Ruark JL, Rodda PR, Morvee WT, VanWitzenberg MJ. Development of chewing in children from 12 to 48 months: Longitudinal study of EMG patterns. *Journal of Neurophysiology*. 1997; 77:2704–2727. [PubMed: 9163386]
- Greenberg S, Carvey H, Hitchcock L, Chang S. Temporal properties of spontaneous speech—A syllable-centric perspective. *Journal of Phonetics*. 2003; 31:465–485.
- Gross CG, Bender DB, Rocha-Miranda CE. Visual receptive fields of neurons in inferotemporal cortex of monkey. *Science*. 1969; 166:1303–1306. [PubMed: 4982685]
- Gross CG, Rocha-Miranda CE, Bender DB. Visual properties of neurons in inferotemporal cortex of macaque. *Journal of Neurophysiology*. 1972; 35:96–111. [PubMed: 4621506]
- Gustison ML, le Roux A, Bergman TJ. Derived vocalizations of geladas (*Theropithecus gelada*) and the evolution of vocal complexity in primates. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences*. 2012; 367:1847–1859.
- Hauser MD, Evans CS, Marler P. The role of articulation in the production of rhesus-monkey, *Macacumulatta*, vocalizations. *Animal Behaviour*. 1993; 45:423–433.
- Hauser MD, Ybarra MS. The role of lip configuration in monkey vocalizations—Experiments using xylocaine as a nerve block. *Brain and Language*. 1994; 46:232–244. [PubMed: 8137144]
- Hiiemae KM, Palmer JB. Tongue movements in feeding and speech. *Critical Reviews in Oral Biology & Medicine*. 2003; 14:413–429. [PubMed: 14656897]
- Hiiemae KM, Palmer JB, Medicis SW, Hegener J, Jackson BS, Lieberman DE. Hyoid and tongue surface movements in speaking and eating. *Archives of Oral Biology*. 2002; 47:11–27. [PubMed: 11743928]
- Hinde RA, Rowell TE. Communication by posture and facial expressions in the rhesus monkey (*Macaca mulatta*). *Proceedings of the Zoological Society London*. 1962; 138:1–21.
- Huber E. Evolution of facial musculature and cutaneous field of trigeminus, Part I. *Quarterly Review of Biology*. 1930a; 5:133–188.

- Huber E. Evolution of facial musculature and cutaneous field of trigeminus, Part II. *Quarterly Review of Biology*. 1930b; 5:389–437.
- Izumi A, Kojima S. Matching vocalizations to vocalizing faces in a chimpanzee (*Pan troglodytes*). *Animal Cognition*. 2004; 7:179–184. [PubMed: 15015035]
- Jarvis ED. Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*. 2004; 1016:749–777. [PubMed: 15313804]
- Jordan KE, Brannon EM, Logothetis NK, Ghazanfar AA. Monkeys match the number of voices they hear with the number of faces they see. *Current Biology*. 2005; 15:1034–1038. [PubMed: 15936274]
- Karmel BZ, Lester ML, McCarvill SL, Brown P, Hofmann MJ. Correlation of infant's brain and behavior response to temporal changes in visual stimulation. *Psychophysiology*. 1977; 14:134–142. [PubMed: 847064]
- Keysers C, Gazzola V. Towards a unifying theory of social cognition. *Progress in Brain Research*. 2006; 156:379–401. [PubMed: 17015092]
- Kiliaridis S, Karlsson S, Kjellberge H. Characteristics of masticatory mandibular movements and velocity in growing individuals and young adults. *Journal of Dental Research*. 1991; 70:1367–1370. [PubMed: 1939831]
- Levitt A, Wang Q. Evidence for language-specific rhythmic influences in the reduplicative babbling of French-and English-learning infants. *Language and Speech*. 1991; 34:235–239. [PubMed: 1843525]
- Lewkowicz DJ. Developmental changes in infants' response to temporal frequency. *Developmental Psychobiology*. 1985; 21:858–865.
- Lieberman DE, McCarthy RC, Hiiemae KM, Palmer JB. Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology*. 2001; 46:117–128. [PubMed: 11163319]
- Lindblom, B.; Krull, D.; Stark, J. Phonetic systems and phonological development. In: de Boysson-Bardies, B.; de Schonen, S.; Jusczyk, P.; MacNeilage, PF.; Morton, J., editors. *Developmental neurocognition: Speech and face processing in the first year of life*. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1996. p. 399-409.
- Locke, JL. *The child's path to spoken language*. Cambridge, MA: Harvard University Press; 1993.
- Locke, JL. Lipsmacking and babbling: Syllables, sociality, and survival. In: Davis, BL.; Zajdo, K., editors. *The syllable in speech production*. New York: Lawrence Erlbaum Associates; 2008. p. 111-129.
- Lund JP, Kolta A. Brainstem circuits that control mastication: Do they have anything to say during speech? *Journal of Communication Disorders*. 2006; 39:381–390. [PubMed: 16884732]
- Lynch M, Oller DK, Steffens M, Buder E. Phrasing in pre-linguistic vocalizations. *Developmental Psychobiology*. 1995; 23:3–25. [PubMed: 7895922]
- MacNeilage PF. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*. 1998; 21:499–511. [PubMed: 10097020]
- MacNeilage, PF. *The origin of speech*. Oxford, UK: Oxford University Press; 2008.
- Malecot A, Johnson R, Kizziar P-A. Syllable rate and utterance length in French. *Phonetica*. 1972; 26:235–251. [PubMed: 4670762]
- Malkova L, Heuer E, Saunders RC. Longitudinal magnetic resonance imaging study of rhesus monkey brain development. *European Journal of Neuroscience*. 2006; 24:3204–3212. [PubMed: 17156381]
- Matsuo K, Palmer JB. Kinematic linkage of the tongue, jaw, and hyoid during eating and speech. *Archives of Oral Biology*. 2010; 55:325–331. [PubMed: 20236625]
- Moore CA, Ruark JL. Does speech emerge from earlier appearing motor behaviors? *Journal of Speech and Hearing Research*. 1996; 39:1034–1047. [PubMed: 8898256]
- Moore CA, Smith A, Ringel RL. Task specific organization of activity in human jaw muscles. *Journal of Speech and Hearing Research*. 1988; 31:670–680. [PubMed: 3230897]
- Morrill RJ, Paukner A, Ferrari PF, Ghazanfar AA. Monkey lip-smacking develops like the human speech rhythm. *Developmental Science*. 2012; 15:557–568. [PubMed: 22709404]
- Nathani S, Oller DK, Cobo-Lewis A. Final syllable lengthening (FSL) in infant vocalizations. *Journal of Child Language*. 2003; 30:3–25. [PubMed: 12718291]

- Newcomb JM, Sakurai A, Lillis JL, Gunaratne CA, Katz PS. Homology and homoplasy of swimming behaviors and neural circuits in the Nudipleura (Mollusca, Gastropoda, Opisthobranchia). *Proceedings of the National Academy of Sciences, U.S.A.* 2012; 109:10669–10676.
- Oller, DK. *The emergence of the speech capacity.* Mahwah, NJ: Lawrence Erlbaum; 2000.
- Ostry DJ, Munhall KG. Control of jaw orientation and position in mastication and speech. *Journal of Neurophysiology.* 1994; 71:1528–1545. [PubMed: 8035233]
- Parr LA. Perceptual biases for multimodal cues in chimpanzee (*Pan troglodytes*) affect recognition. *Animal Cognition.* 2004; 7:171–178. [PubMed: 14997361]
- Parr LA, Cohen M, de Waal F. Influence of social context on the use of blended and graded facial displays in chimpanzees. *International Journal of Primatology.* 2005; 26:73–103.
- Preuschoff K, Quartz SR, Bossaerts P. Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience.* 2008; 28:2745–2752. [PubMed: 18337404]
- Redican, WK. Facial expressions in nonhuman primates. In: Rosenblum, LA., editor. *Primate behavior: Developments in field and laboratory research.* New York: Academic Press; 1975. p. 103–194.
- Ross CF, Reed DA, Washington RL, Eckhardt A, Anapol F, Shahnoor N. Scaling of chew cycle duration in primates. *American Journal of Physical Anthropology.* 2009; 138:30–44. [PubMed: 18661474]
- Saberi K, Perrott DR. Cognitive restoration of reversed speech. *Nature.* 1999; 398:760. [PubMed: 10235257]
- Schneirla, TC. Levels in the psychological capacities of animals. In: Sellars, RW.; McGill, VJ.; Farber, M., editors. *Philosophy for the future.* New York: Macmillan; 1949. p. 243–286.
- Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science.* 1995; 270:303–304. [PubMed: 7569981]
- Shepherd SV, Lanzilotto M, Ghazanfar AA. Facial muscle coordination during rhythmic facial expression and ingestive movement. *Journal of Neuroscience.* 2012; 32:6105–6116. [PubMed: 22553017]
- Sherwood CC. Comparative anatomy of the facial motor nucleus in mammals, with an analysis of neuron numbers in primates. *Anatomical Record Part A—Discoveries in Molecular Cellular and Evolutionary Biology.* 2005; 287A:1067–1079.
- Sherwood CC, Hof PR, Holloway RL, Semendeferi K, Gannon PJ, Frahm HD, et al. Evolution of the brainstem orofacial motor system in primates: A comparative study of trigeminal, facial, and hypoglossal nuclei. *Journal of Human Evolution.* 2005; 48:45–84. [PubMed: 15656936]
- Sherwood CC, Holloway RL, Erwin JM, Hof PR. Cortical orofacial motor representation in old world monkeys, great apes, and humans—II. Stereologic analysis of chemoarchitecture. *Brain Behavior and Evolution.* 2004; 63:82–106.
- Sherwood CC, Holloway RL, Erwin JM, Schleicher A, Zilles K, Hof PR. Cortical orofacial motor representation in old world monkeys, great apes, and humans—I. Quantitative analysis of cytoarchitecture. *Brain Behavior and Evolution.* 2004; 63:61–81.
- Sliwa J, Duhamel JR, Pascalis O, Wirth S. Spontaneous voice–face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences, U.S.A.* 2011; 108:1735–1740.
- Smith A, Zelaznik HN. Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology.* 2004; 45:22–33. [PubMed: 15229873]
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature.* 2002; 416:87–90. [PubMed: 11882898]
- Steckenfinger SA, Ghazanfar AA. Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences, U.S.A.* 2009; 106:18362–18466.
- Steeve RW. Babbling and chewing: Jaw kinematics from 8 to 22 months. *Journal of Phonetics.* 2010; 38:445–458. [PubMed: 20725590]
- Steeve RW, Moore CA, Green JR, Reilly KJ, McMurtrey JR. Babbling, chewing, and sucking: Oromandibular coordination at 9 months. *Journal of Speech Language and Hearing Research.* 2008; 51:1390–1404.

- Sugihara T, Diltz MD, Averbeck BB, Romanski LM. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *Journal of Neuroscience*. 2006; 26:11138–11147. [PubMed: 17065454]
- Takahashi DY, Baccala LA, Sameshima K. Information theoretic interpretation of frequency domain connectivity measures. *Biological Cybernetics*. 2010; 103:463–469. [PubMed: 21153835]
- Thelen E. Rhythmical behavior in infancy—An ethological perspective. *Developmental Psychology*. 1981; 17:237–257.
- Van Hooff JARAM. Facial expressions of higher primates. *Symposium of the Zoological Society, London*. 1962; 8:97–125.
- van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, U.S.A.* 2005; 102:1181–1186.
- Vitkovitch M, Barber P. Visible speech as a function of image quality: Effects of display parameters on lipreading ability. *Applied Cognitive Psychology*. 1996; 10:121–140.
- von Kriegstein K, Dogan O, Gruter M, Giraud AL, Kell CA, Gruter T, et al. Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences, U.S.A.* 2008; 105:6747–6752.
- Yehia H, Kuratate T, Vatikiotis-Bateson E. Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*. 2002; 30:555–568.
- Yehia H, Rubin P, Vatikiotis-Bateson E. Quantitative association of vocal-tract and facial behavior. *Speech Communication*. 1998; 26:23–43.
- Zangenehpour S, Ghazanfar AA, Lewkowicz DJ, Zatorre RJ. Heterochrony and cross-species intersensory matching by infant vervet monkeys. *PLoS One*. 2009; 4:e4302. [PubMed: 19172998]

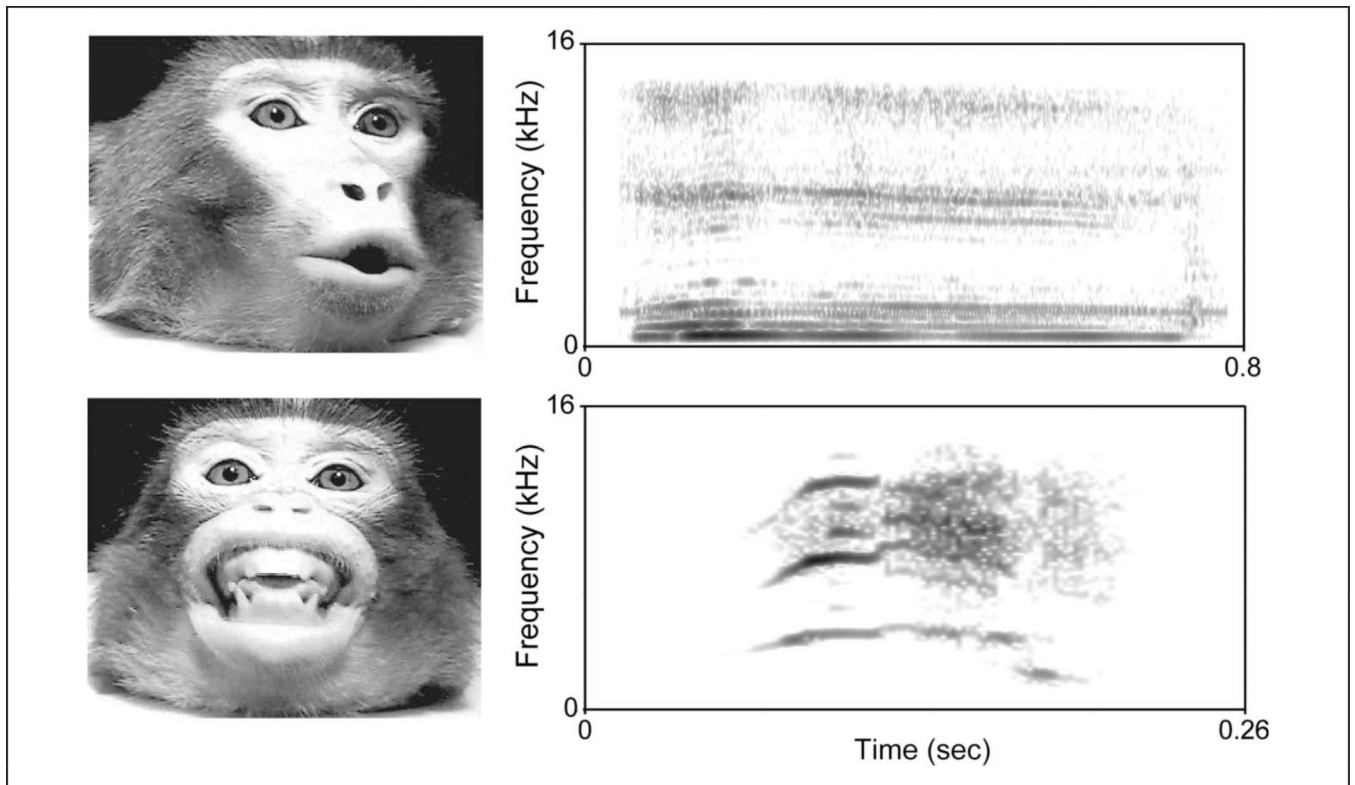


Figure 1. Different facial expressions are produced concomitantly with different vocalizations. Rhesus monkey coo and scream calls. Video frames extracted at the midpoint of the expressions with their corresponding spectrograms. X axis depicts time in seconds; y axis depicts frequency in kHz.

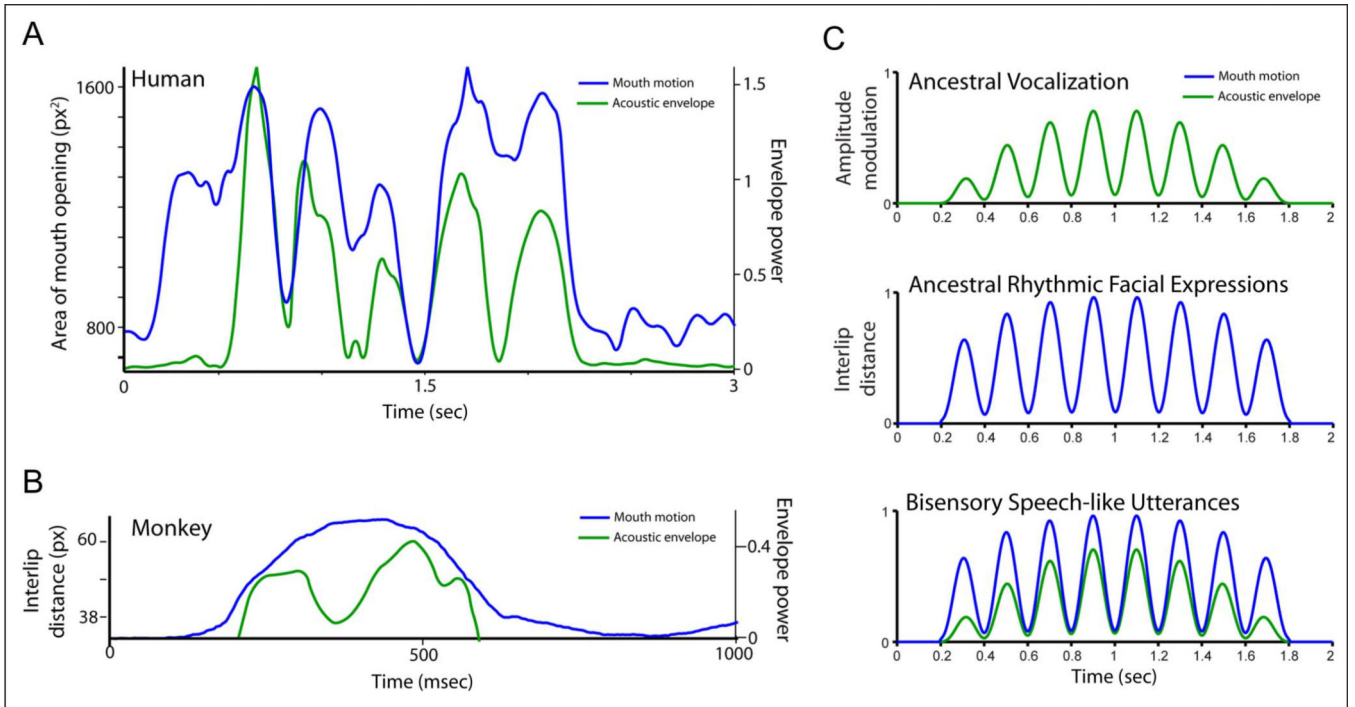


Figure 2.

Rhythmic structure of vocal signals in humans and monkeys. (A) Mouth motion and auditory envelope for a single sentence produced by human. The *x* axis depicts time in seconds; the *y* axis on the left depicts the area of the mouth opening in pixel squared; and the *y* axis on the right depicts the acoustic envelope in Hilbert units. (B) Mouth motion and the auditory envelope for a single coo vocalization produced by a macaque monkey. The *x* axis depicts time in milliseconds; the *y* axis on the left depicts the distance between lips in pixels; and the *y* axis on the right depicts the acoustic envelope power in Hilbert units. (C) A version of MacNeilage’s hypothesis for the evolution of rhythmic speech.

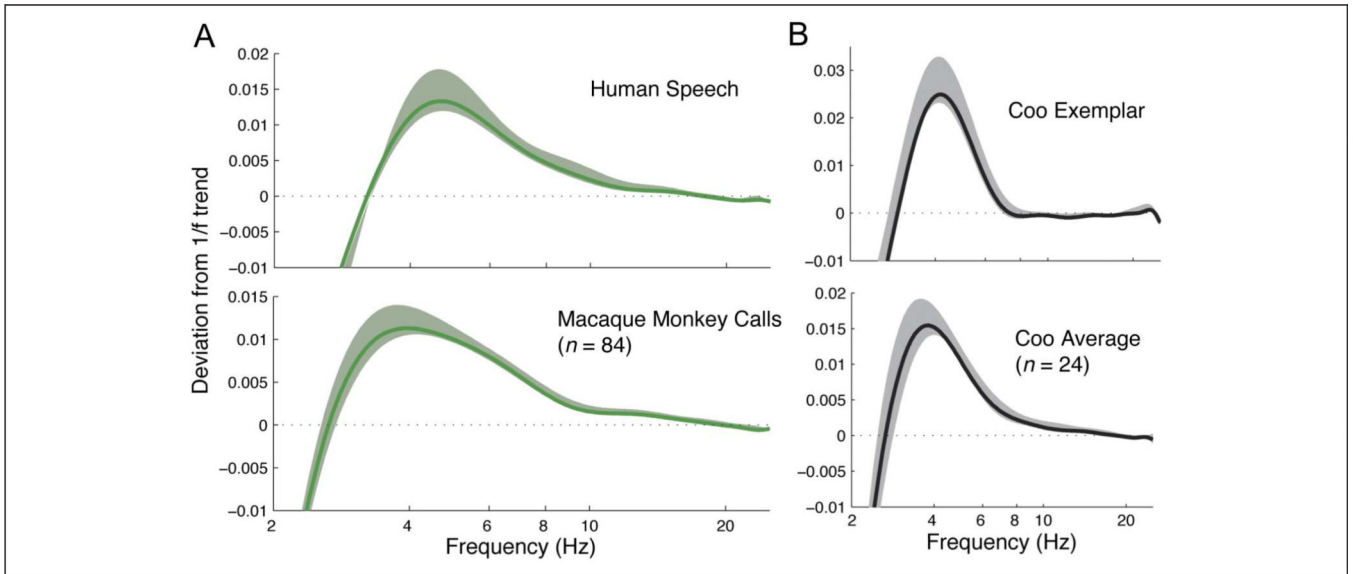


Figure 3.

Speech and macaque monkey calls have similar rhythmic structure in their acoustic envelopes. (A) Modulation spectra for human speech and long-duration (>400 msec) macaque monkey calls. (B) Modulation spectra for coo calls and an exemplar of a coo call, respectively. The x axes depict power deviations from a $1/f$ trend; the y axes represent frequency in log Hz.

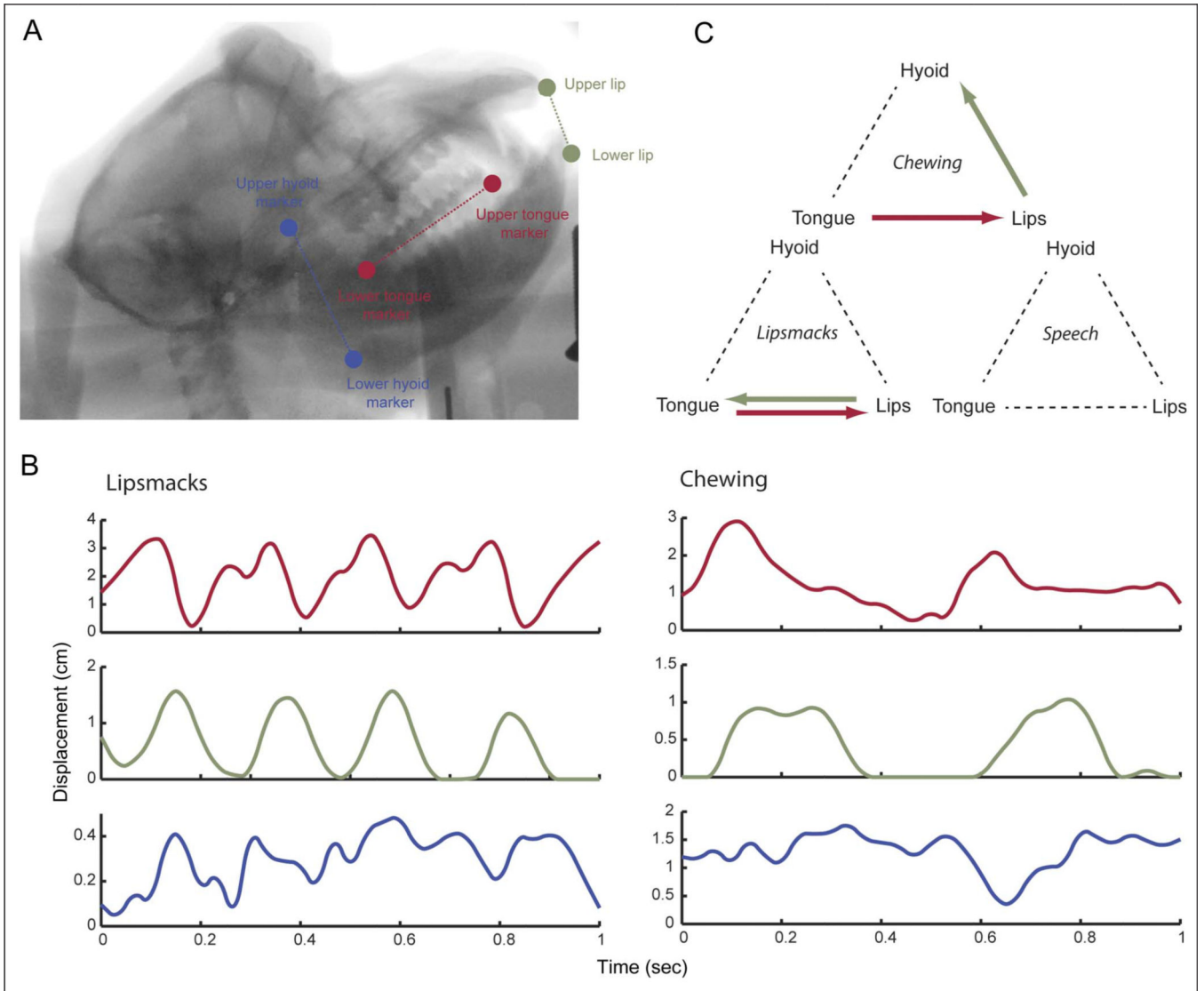


Figure 4. Internal biomechanics of rhythmic orofacial movements. (A) The anatomy of the macaque monkey vocal tract as imaged with cineradiography. The key vocal tract structures are labeled: the lips, tongue, and hyoid. (B) Time-displacement plot of the tongue, interlip distance, and hyoid for one exemplar each of lip smacking and chewing. (C) Arrow schematics show the direction of significant influence from each structure onto the other two as measured by the partial directed coherence analysis of signals such as those in B. Modified with permission from Ghazanfar et al. (2012).

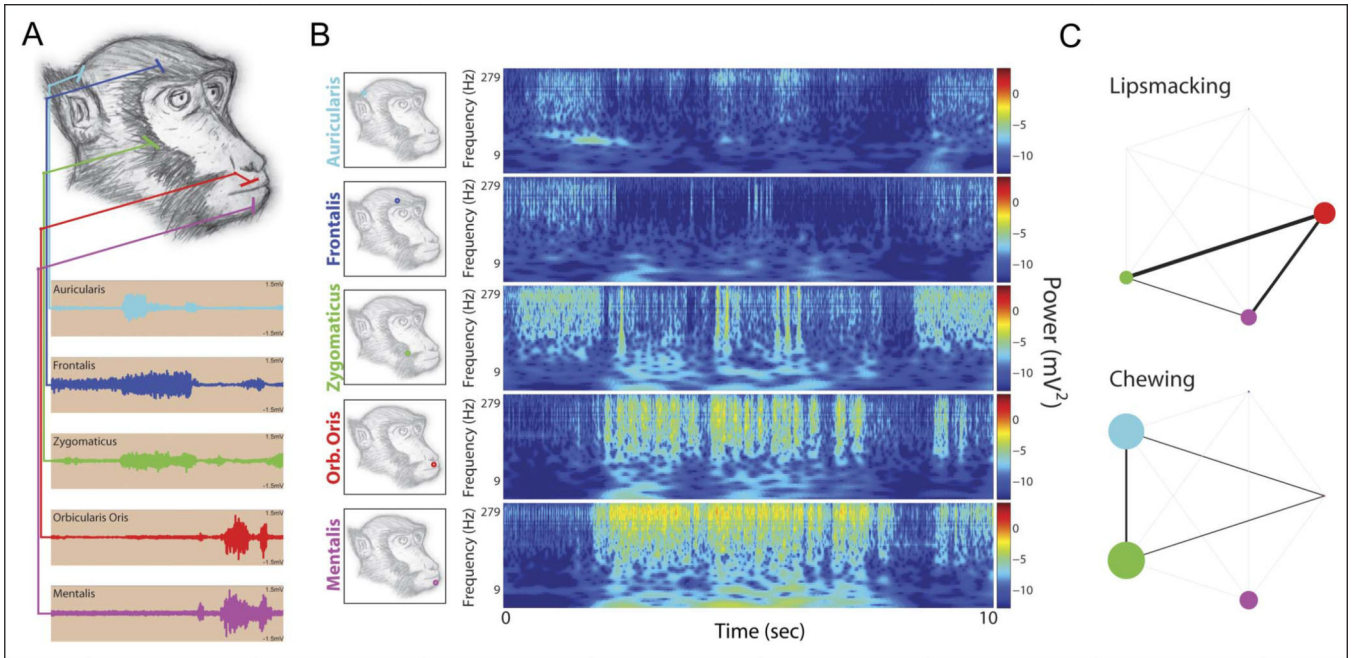


Figure 5.

Facial electromyography during lip smacking and chewing. (A) Acute indwelling electrodes were inserted into the auricularis (cyan), frontalis (blue), zygomaticus (green), orbicularis oris (red), and mentalis (magenta) facial muscles. These muscles contribute to facial expression. Each recording site tapped independent electrical activities that corresponded with video-monitored muscle tension. (B) Example of lip-smacking bout. Lip smacking was characterized by rhythmic activity in the orofacial muscles. Although upper facial muscles controlling the ear and brow have been implicated in natural observations, they appeared to play an intermittent and nonrhythmic role, suggesting independent control. X axis = time in seconds; y axis = log frequency in Hz. (C) Muscle rhythm coordination as measured by coherence. Significant power modulations and modulation coherencies are depicted for each of the muscle groups (auricularis = cyan; frontalis = blue; zygomaticus = green; orbicularis oris = red; mentalis = magenta). Node weight corresponds to the total amount by which measured power modulations exceeded a permutation baseline; line weight corresponds to the total amount by which measured coherency exceeded the permutation baseline. Orb. = orbicularis. Reprinted with permission from Shepherd et al. (2012).

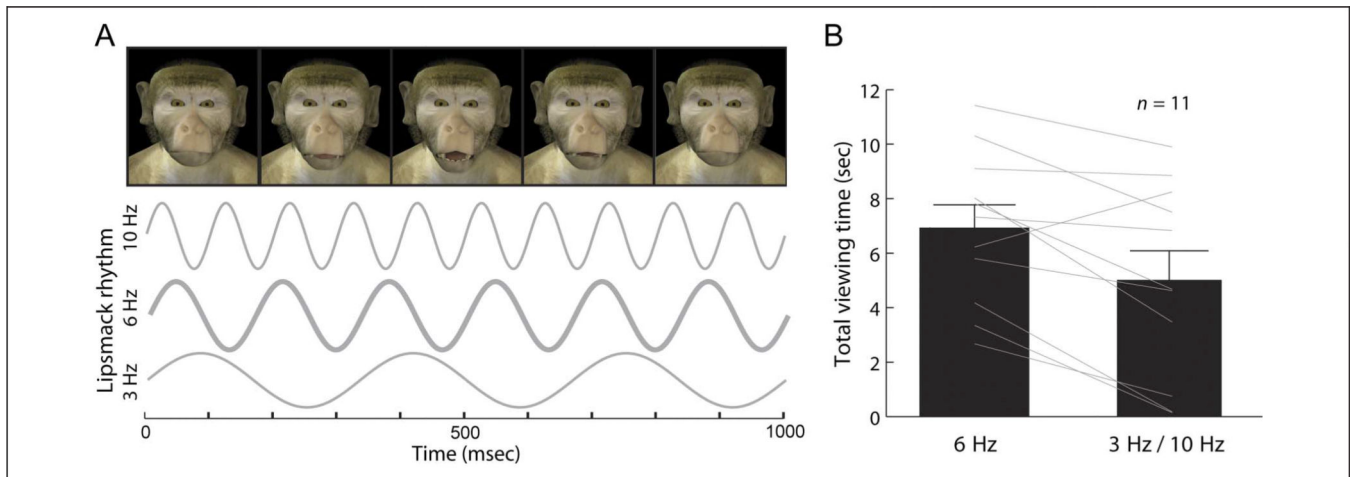


Figure 6.

Perception of lip smacking is tuned to the 3- to 8-Hz rhythm. (A) Synthetic lip-smacking rates were presented faster (10 Hz) or slower (3 Hz) than the natural rate (6 Hz). (B) Total viewing times in seconds for individual subjects (lines) and grand total (mean and standard error). All but one subject showed a preference for the avatar with species-typical lip-smacking rate. Reprinted with permission from Ghazanfar et al. (2013).