

Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm

Asif A. Ghazanfar^{a,b,c,1}, Ryan J. Morrill^{a,c}, and Christoph Kayser^{d,e,1}

^aNeuroscience Institute, Departments of ^bPsychology and ^cEcology and Evolutionary Biology, Princeton University, Princeton, NJ 08540; ^dMax Planck Institute for Biological Cybernetics, 72076 Tuebingen, Germany; ^eInstitute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, United Kingdom

Edited* by Charles Gross, Princeton University, Princeton, NJ, and approved December 5, 2012 (received for review August 29, 2012)

Human speech universally exhibits a 3- to 8-Hz rhythm, corresponding to the rate of syllable production, which is reflected in both the sound envelope and the visual mouth movements. Artificial perturbation of the speech rhythm outside the natural range reduces speech intelligibility, demonstrating a perceptual tuning to this frequency band. One theory posits that the mouth movements at the core of this speech rhythm evolved through modification of ancestral primate facial expressions. Recent evidence shows that one such communicative gesture in macaque monkeys, lip-smacking, has motor parallels with speech in its rhythmicity, its developmental trajectory, and the coordination of vocal tract structures. Whether monkeys also exhibit a perceptual tuning to the natural rhythms of lip-smacking is unknown. To investigate this, we tested rhesus monkeys in a preferential-looking procedure, measuring the time spent looking at each of two side-by-side computer-generated monkey avatars lip-smacking at natural versus sped-up or slowed-down rhythms. Monkeys showed an overall preference for the natural rhythm compared with the perturbed rhythms. This lends behavioral support for the hypothesis that perceptual processes in monkeys are similarly tuned to the natural frequencies of communication signals as they are in humans. Our data provide perceptual evidence for the theory that speech may have evolved from ancestral primate rhythmic facial expressions.

One universal feature of speech is its rhythm. 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 (1–4). This 3- to 8-Hz rhythm is critical to speech perception: Disrupting the auditory component of this rhythm significantly reduces intelligibility (5–9), as does disrupting the visual dynamics arising from mouth and facial movements (10). The exquisite sensitivity of speech perception to this rhythm is thought to be related to on-going neural rhythms in the neocortex. Rhythmic activity in the auditory cortex prevails in a similar 3- to 8-Hz (theta) frequency range (11–13), and the temporal signature of the neural rhythm (i.e., its phase) locks to the temporal dynamics of complex sounds such as speech (14–17). Importantly, speeding up speech so that it is faster than 8 Hz disrupts the ability of the auditory cortex to track speech (14, 16), whereas adding silent intervals to time-compressed speech can rescue comprehension when the interval frequency rate is matched to the theta range (18). Thus, the natural rhythm of speech seems to be linked to on-going neocortical oscillations.

Given the importance of this rhythm in speech, understanding how speech evolved requires investigating the origins of its rhythmic structure and the brain's sensitivity to it. One theory posits that the rhythm of speech evolved through the modification of rhythmic facial movements in ancestral primates (19). Such facial movements are extremely common in the form of visual communicative gestures. Lip-smacking, for example, is an affiliative signal observed in many genera of primates (20–22), including chimpanzees (23). It is characterized by regular cycles of vertical jaw movement, often involving a parting of the lips, but sometimes occurring with closed, puckered lips (24). Importantly, as a communication signal, lip-smacking is directed at another individual during face-to-face interactions (22, 25) and is among

the first facial expressions produced by infant monkeys (26, 27). According to MacNeilage (19), 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.

Although direct tests of such evolutionary hypotheses are difficult, we recently showed that the production of lip-smacking in macaque monkeys is, indeed, strikingly similar (likely homologous) to the orofacial rhythms produced during speech. For example, in contrast to chewing and other rhythmic orofacial movements, lip-smacking exhibits a speech/theta-like 3- to 8-Hz rhythm (24, 28, 29), and its developmental trajectory is the same as the trajectory leading from human babbling to adult consonant-vowel utterances (29). Moreover, an X-ray cineradiographic study of the dynamics of vocal tract elements (lips, tongue, and hyoid bone) during lip-smacking versus chewing showed that the differential functional coordination of these effectors during these behaviors parallels that of human speech and chewing (24). However, what about perception? At the neural level, we know that, as in humans, temporal lobe structures in the macaque monkey (e.g., auditory cortex, superior temporal sulcus) are sensitive to dynamic audiovisual vocal communication signals (28, 30–33) and that they exhibit rhythmic neural activity fluctuations in the theta range spontaneously (34) and in response to naturalistic stimuli (35). However, we do not know if monkeys are perceptually tuned to the species-typical rhythmicity in the same way that humans are perceptually most sensitive to natural speech rhythms.

Results

We used a preferential-looking procedure to test whether monkeys were differentially sensitive to lip-smacking produced with a rhythmic frequency in the species-typical range (mean: 4–6 Hz) (24, 28, 29). Because we cannot ask monkeys to produce faster and slower versions of their facial expressions, we used computer-generated monkey avatars to produce stimuli varying in lip-smacking frequency but with otherwise identical features (Fig. 1A) (36, 37). The use of avatar faces allowed us to control additional factors, such as head and eye movements and lighting conditions for face and background, that could potentially influence looking times. Each of two unique avatar faces was generated to produce three different lip-smacking rhythms: 3, 6, and 10 Hz (Fig. 1B). In every test session, one avatar was lip-smacking at 6 Hz whereas the other was lip-smacking at either 3 or 10 Hz. The use of the two avatar identities and their side of presentation on the screen was counterbalanced across subjects.

We assessed preferential looking by measuring looking times to one or the other avatar while these were presented on a wide

Author contributions: A.A.G. and R.J.M. designed research; C.K. performed research; A.A.G. contributed new reagents/analytic tools; A.A.G., R.J.M., and C.K. analyzed data; and A.A.G. and C.K. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

¹To whom correspondence may be addressed. E-mail: asifg@princeton.edu or kayser@tuebingen.mpg.de.

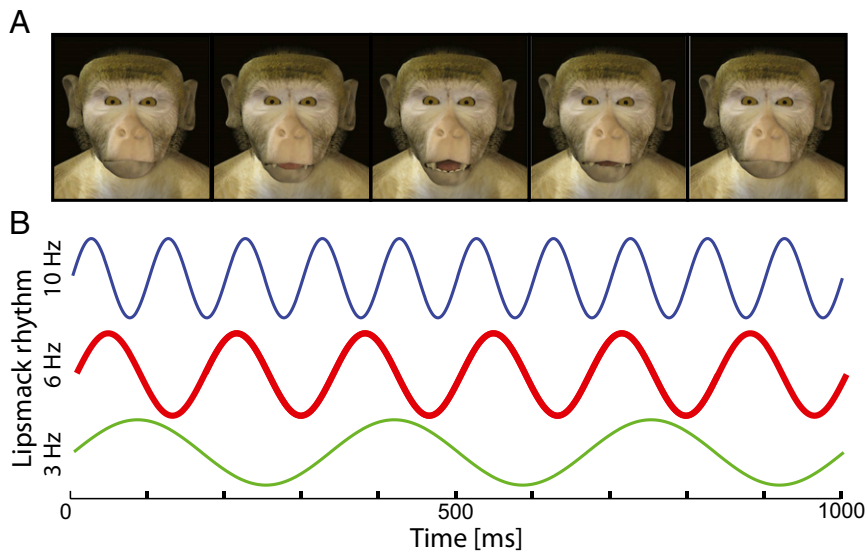


Fig. 1. Example of a frame sequence from a video clip showing an avatar face producing a lip-smacking gesture. Lip-smacking is characterized by regular cycles of vertical jaw movement, often involving a parting of the lips.

screen in front of the subject. 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 ($3 > 6 > 10$ Hz). Third, they could prefer faster rhythms ($3 < 6 < 10$ Hz) (38). Fourth, they could show avoidance of the 6-Hz lip-smacking, preferring the unnatural 3- and 10-Hz rhythms over the natural lip-smacking rhythm. This may arise if monkeys find the naturalistic 6-Hz lip-smacking disturbing [perhaps uncanny (37)] or too arousing (39). 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 (13, 40).

We analyzed behavioral data from 11 monkey subjects; each subject was tested only once to avoid habituation (41–44). Monkeys showed a preference for the 6-Hz lip-smacking rhythm. Total viewing times were longer for the 6-Hz avatar (6.9 ± 0.84 s, mean \pm SEM) compared with the 3- and 10-Hz avatars (5.0 ± 1.0 s), and this difference was statistically significant (paired *t* test, $n = 11$, $P < 0.01$) (Fig. 2A). All but 1 of the 11 subjects

expressed a viewing preference for the 6-Hz versus the 3- or 10-Hz avatars (binomial test, $P = 0.006$). When expressed as relative viewing-time preference, the monkeys on average spent $30 \pm 11\%$ more time looking at the 6-Hz versus the other avatars (Fig. 2B). A similar preference for the 6-Hz avatar was evident compared with the either slower or faster lip-smacking avatars separately. Subjects tested on 6- vs. 3-Hz and 6- vs. 10-Hz comparisons separately had total viewing times of 6.6 ± 0.84 s vs. 5.3 ± 1.5 s (6 vs. 3 Hz, $n = 5$) and 7.1 ± 1.5 s vs. 4.8 ± 1.6 s (6 vs. 10 Hz, $n = 6$). Of the different potential outcomes, our results provide clear evidence for a perceptual preference toward theta- and speech-like natural frequency of lip-smacking displays. The magnitude of the effect was smaller in the 6- vs. 3-Hz condition perhaps because the 3-Hz rhythm still falls within the species-typical range of lip-smacking rhythms, but 10 Hz does not (24, 28, 29).

There is a possibility that this pattern of preferential looking has little or nothing to do with the faces or lip-smacking per se, but rather with just the rhythmic motion alone; that is, the subjects may not have seen the avatars as conspecific faces producing a species-typical expression. In the wild, lip-smacking gestures are sometimes exchanged: an individual lip-smacking toward another will receive lip-smacking in return. We capitalized on this fact and

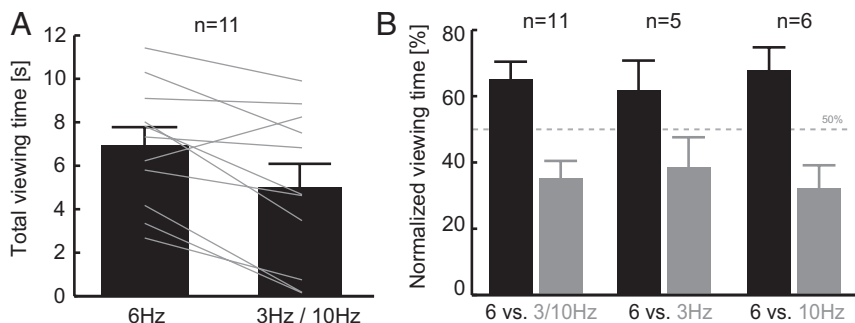


Fig. 2. Preferential viewing times are longer for biological lip-smacking rates (6 Hz) than for slower or faster manipulations (3 or 10 Hz). (A) Total viewing times in seconds for individual subjects (lines) and the grand total (mean \pm SE). All but one subject showed a preference for the avatar with biological lip smack rate. (B) Preferential-looking time for each avatar (i.e., frequency), expressed in percentage of the total time viewing either avatar. Bars indicate the mean \pm SE across subjects for all comparisons (6 vs. 3 or 10 Hz) and individually for subject groups tested on 6 vs. 3 Hz and 6 vs. 10 Hz, respectively. Subjects consistently preferred the 6-Hz lip-smacking versus other frequencies.

re-examined the videos of our subjects to see whether, beyond their looking preferences, they exhibited lip-smacking toward the avatars. Remarkably, the 5 of the 11 subjects produced their own lip-smacking gestures in response to the avatar faces. This proportion (~45%) is roughly the same as that seen in the wild (29). Fig. 3 shows examples of movement trajectories from three of those lip-smacking subjects. Note that the rhythmic frequency of these lip-smacking responses falls well within the normal range for both captive (28) and wild (29) macaque monkeys. The lip-smacking responses by our subjects demonstrate that the avatars and their lip-smacking motion were viewed as “real” facial expressions.

Discussion

Our data show that monkeys are perceptually tuned to lip-smacking expressions within the species-typical range around 6 Hz (24, 28, 29). This is analogous, and likely homologous (see below), to the tuning of human perception to natural speech rhythms. Speech fluctuates in the 3- to 8-Hz band, corresponding to the rate of syllable production (1–4). If this rhythm is eliminated or sped up beyond 8 Hz, speech intelligibility declines precipitously (5–9, 18). Importantly, the rhythmic nature of speech also applies to the visual domain, as facial and mouth movements have a similar rhythmicity that is tightly correlated with the auditory stream (1).

The neural basis of this perceptual sensitivity may arise from a match of the rhythms in speech to on-going theta activity in the neocortex (45). Activity in the auditory cortex phase-locks to the temporal structure of speech, and this ability to track and decode the acoustic input is reduced when speech is played back at faster than normal rates (14, 16, 46). Even visual displays of speech can entrain neural activity in the auditory cortex (15), thereby enhancing the encoding of auditory information (47). Theta rhythms are ubiquitous in the neocortices of mammals (48), including the temporal lobe structures of macaque monkeys (30, 34) where rhythmic activity was found to be sensitive to dynamic facial expressions (28, 30–32), as observed in humans. Thus, as it is for speech in the human brain, the neural substrates are similarly present in the macaque monkeys to potentially undergird their perceptual sensitivity to natural lip-smacking rhythm as observed here (28). It is also possible that the perceptual tuning to the natural rhythms of speech and lip-smacking is simply due to the abundant experience that humans and monkeys, respectively, have with these social signals. Although this scenario does not preclude a role for neocortical rhythms, it also does not require them. Moreover, it is possible that neocortical rhythms

are themselves “tuned” by these social signals. Establishing a direct causal role for neocortical rhythms in communication is one of the big challenges for future work.

Beyond the domain of communication, developmental studies of human infants show perceptual and neural tuning to the temporal frequency of visual patterns (38, 40). In a preferential-looking procedure similar to the one we adopted, 4-mo-old infants were presented with a checkerboard pattern flashed at frequencies from 1 to 20 Hz (40). Infants showed a perceptual preference for frequencies around 5–6 Hz, and visually evoked potentials measured in the same-aged infants exhibited maximal amplitudes when presented with stimuli flashing at the same temporal frequencies (40). This mimics the tight link proposed between the rhythms in speech and neural activity described above. Indeed, Karmel and Maisel (49) suggested that the looking preference of infants (i.e., the strength of their orienting response) is linked to integration of synchronous neural activity with the temporal frequency of the sensory signal. The same processes are likely relevant for the preferential looking of adult monkeys toward natural lip-smacking expressions. However, the fact that checkerboards can elicit the same rhythmic preference in human infants suggests that the perceptual preference to theta-range stimuli is more generic and extends beyond facial expressions. We could not test this possibility because adult monkeys usually do not attend sufficiently long to behaviorally irrelevant stimuli to obtain meaningful data in preferential-looking procedures (44). Nevertheless, it is quite possible that stimulus identity matters a great deal in tests of temporal frequency sensitivity and its development (38).

The tight link between the brain’s rhythmic organization and the rhythmic patterns of communication signals also extends beyond the orofacial domain. Monkeys and apes produce non-vocal acoustic gestures by “drumming” with their hands on other body parts or objects or by rhythmically shaking an object in their environment (50–52). Macaque monkeys frequently produce such rhythmic gestures by shaking the branches of a tree or by hitting the ground, resulting in periodic sounds. Captive macaques exhibit a similar behavior by shaking environmental objects (53). Intriguingly, the rhythmic structure of this nonfacial acoustic gesture falls into the theta range as well, with typical frequencies around 5 Hz. Using preferential-orienting tests, the behavioral relevance of such drumming sounds was demonstrated. Moreover, these acoustic gestures preferentially activate those auditory cortical regions known to be sensitive to vocal communication sounds (53), which may generate the prominent low-frequency rhythms observed in the auditory cortex (17).

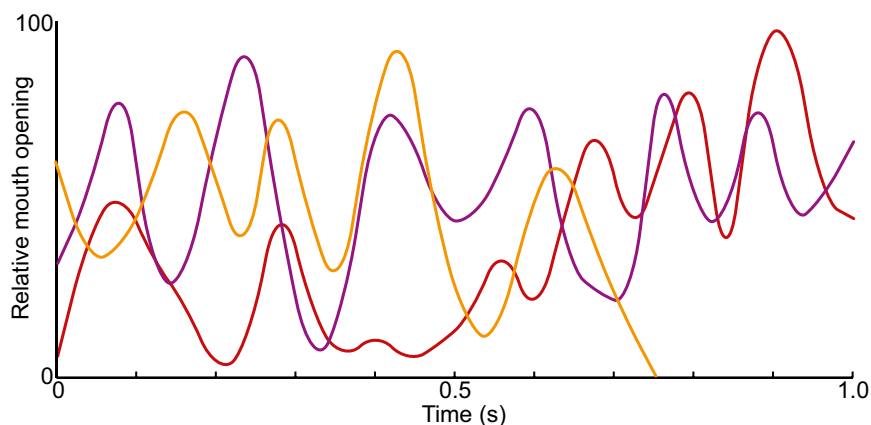


Fig. 3. Monkey subjects respond to lip-smacking avatars with their own lip-smacking facial expressions. Three exemplars of lip-smacking movement trajectories from subjects produced when they viewed lip-smacking avatars. The y axis represents normalized units for mouth opening, with peak interlip distance during the sequence set at 100; the x axis is time in seconds.

Thus, communicative audiovisual displays, whether produced using hands or orofacial structures, exhibit similar temporal organization, suggesting that the link between the rhythmic organization of cortical activity and the rhythmic structure of communication signals is general and effector-independent.

Our finding that monkeys display a similar perceptual sensitivity to lip-smacking rhythmicity as humans do for speech supports an influential theory by MacNeilage, which suggests that speech evolved through the modification of rhythmic facial movements in ancestral primates (19). In the domain of orofacial motor control, speech movements are faster than chewing movements (54–58), and an X-ray cineradiography study revealed that, during lip-smacking, the lips, tongue, and hyoid are loosely coordinated as in speech and move with a theta-like rhythm (24). Chewing movements, in comparison, were significantly slower and differently coordinated when compared with lip-smacking (as it is for human chewing versus speech). Thus, the production of lip-smacking and speech is strikingly similar at the level of motor control. Along the same lines, the developmental trajectory of monkey lip-smacking parallels speech development (29, 59). Measurements of the rhythmicity and variability of lip-smacking across age groups (neonates, juveniles, and adults) revealed that movement variability decreases with age whereas speed increases: this is exactly as speech develops—from babbling to adult consonant-vowel production (29). Importantly, the developmental trajectory for lip-smacking was different from that of chewing (29), in line with differences in the developmental trajectories between speech and chewing reported for humans (60–62). Combined with the perceptual tuning to natural lip-smacking rhythms reported here, these data suggest that monkey lip-smacking and human speech rhythms share a homologous sensorimotor mechanism. These results lend strong empirical support for the idea that human speech evolved from the rhythmic facial expressions or other rhythmic actions normally produced by our primate ancestors (19).

Methods

Subjects. Behavioral data were obtained from 12 adult rhesus monkeys (*Macaca mulatta*) that were part of a larger colony group housed in enriched environments. All procedures were approved by local authorities (Regierungspräsidium Tübingen), were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC), and were in concordance with the recommendations of the Weatherall report on the use of nonhuman primates in research. As in our previous preferential-looking studies with monkeys, each subject was tested only once to avoid familiarity with the stimuli and because subjects quickly habituate to such testing environments (41–44). Of the 12 subjects tested, 1 did not behave calmly in the experiment booth and thus did not consistently look at either side of the screen; this subject was not included in our analyses. Such exclusions are typical of preferential-looking studies using human infants.

Preferential Looking Procedure. The preferential-looking stimuli consisted of two computer-generated synthetic monkeys (avatars) on opposite sides of the screen, lip-smacking at parametrically varied rates. These realistic synthetic monkey faces allowed for manipulation of the temporal dynamics of mouth oscillation independently of other visual properties of the dynamic face, such as eye and head movements, which might confound looking time if

temporally manipulated. The mean natural lip-smacking frequency for macaque monkeys is ~5–6 Hz (24, 28, 29). For this reason, we compared preferential-looking times to avatars' lip-smacking at 6 Hz versus avatars' lip-smacking at 3 or 10 Hz. The slow frequency of 3 Hz was chosen because it is within the range of natural rhythmic facial gestures, but on the edge of the adult lip-smacking range. The 10-Hz frequency was chosen as a faster-than-natural rhythm. During each session the 6- and 3/10-Hz stimuli were produced by a different avatar face to avoid confounds arising from viewing identical monkeys side by side. Also, the assignment of avatars and lip-smacking frequencies and their side presentation on the screen were counterbalanced across subjects to control for individual identity and lateralization preferences as a potential confound. There were no statistically significant patterns of looking related to avatar identity (identity 1 vs. 2: 6.1 ± 1.0 s vs. 5.9 ± 1.0 s, t test $P = 0.8$) or side of presentation (left vs. right: 5.9 ± 0.9 s vs. 6.1 ± 1 s, $P = 0.8$).

Animals were seated in a custom-made and individualized primate chair, which was positioned in a dark and an-echoic booth in front of a large SHARP LCD-TV monitor at eye level with the subject (about 60 cm to the animal's head, 117-cm screen diagonal). The monitor was initially blank, and subjects were given 4–8 min to habituate to the setup. The stimulus videos showed two avatars on the left and right side, with each avatar's face spanning about 15°–20° of visual angle and a distance of about 65° between the centers of both avatar heads. The experimenter monitored the subjects from outside the room using an infrared camera and initiated the start of the stimulus presentation following a period during which the subject was looking straight-ahead toward the monitor. The stimulus video was presented for 60 s, subsequent to which the monitor was blanked and the subject returned to the animal colony. Behavioral data were recorded as digital video files showing as high-resolution (720 × 576 pixels, 30 Hz) close-up images of the subject's head using an infrared camera and WinTV2K Hauppauge software.

Video Scoring. Looking directions were scored manually and by two observers blind to the experimental conditions. For analysis, the 60-s stimulus presentation period was extracted from the video, and for each frame the observer scored whether the subject looked at either of the two avatars or away from them. The total duration of the subjects looking time toward each avatar was recorded and expressed either as total time in seconds or as a fraction of the total time that the subject spent looking at either avatar. Scoring which of the two avatars that the monkeys were looking at was unambiguous, as in previous preferential-looking studies using a very similar experimental setup (41–43). The two avatar faces were far apart in the horizontal dimension, fairly close to the subject's eyes and at eye level. The subject thus had to make clear head or very large eye movements to switch focus from one avatar to another. To validate the scoring procedure, we measured interobserver reliability (Pearson correlation, $r = 0.99$).

Measuring Lip-smacking Production. Five of the 11 subjects responded to the avatars with their own lip-smacking gestures. We measured the trajectories of these movements with a procedure that we have used previously (28, 29). Mouth displacement was measured frame by frame. Displacement was measured by manually indicating one point in the middle of the top lip and one in the middle of the bottom lip for each frame. For some lip-smacking, the top and bottom lips do not part, so interlip distance does not provide a reliable indication of jaw displacement. For these lip-smackings, displacement was measured as the distance between the lower lip and the nasion (the point between the eyes where the bridge of the nose begins), an easily identifiable point that does not move during the gestures.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health (National Institute of Neurological Disorders and Stroke) Grant R01NS054898 (to A.A.G.) and by the Max Planck Society (C.K.).

- Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. *PLoS Comput Biol* 5(7):e1000436.
- Greenberg S, Carvey H, Hitchcock L, Chang S (2003) Temporal properties of spontaneous speech: A syllable-centric perspective. *J Phonetics* 31(3–4): 465–485.
- Crystal TH, House AS (1982) Segmental durations in connected speech signals: Preliminary results. *J Acoust Soc Am* 72(3):705–716.
- Malécot A, Johnston R, Kizziar P-A (1972) Syllabic rate and utterance length in French. *Phonetica* 26(4):235–251.
- Drullman R, Festen JM, Plomp R (1994) Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am* 95(5, Pt 1):2670–2680.
- Saberi K, Perrott DR (1999) Cognitive restoration of reversed speech. *Nature* 398(6730): 760–760.
- Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270(5234):303–304.
- Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416(6876):87–90.
- Elliott TM, Theunissen FE (2009) The modulation transfer function for speech intelligibility. *PLoS Comput Biol* 5(3):e1000302.
- Vitkovitch M, Barber P (1996) Visible speech as a function of image quality: Effects of display parameters on lipreading ability. *Appl Cogn Psychol* 10(2): 121–140.

11. Poeppel D (2003) The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time' *Speech Commun* 41(1): 245–255.
12. Ghitza O (2011) Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Front Psychol* 2:130.
13. Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat Neurosci* 15(4):511–517.
14. Ahissar E, et al. (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci USA* 98(23):13367–13372.
15. Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol* 8(8): e1000445.
16. Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54(6):1001–1010.
17. Ng BS, Logothetis NK, Kayser C (2012) EEG phase patterns reflect the selectivity of neural firing. *Cereb Cortex*, 10.1093/cercor/bhs031.
18. Ghitza O, Greenberg S (2009) On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66(1–2):113–126.
19. MacNeilage PF (1998) The frame/content theory of evolution of speech production. *Behav Brain Sci* 21(4):499–511; discussion 511–546.
20. Hinde RA, Rowell TE (1962) Communication by posture and facial expressions in the rhesus monkey (*Macaca mulatta*). *Proc Zool Soc Lond* 138(1):1–21.
21. Redican WK (1975) Facial expressions in nonhuman primates. *Primate Behavior: Developments in Field and Laboratory Research*, ed Rosenblum LA (Academic Press, New York), pp 103–194.
22. Van Hooff JARAM (1962) Facial expressions of higher primates. *Symp Zoolog Soc Lond* 8:97–125.
23. Parr LA, Cohen M, de Waal F (2005) Influence of social context on the use of blended and graded facial displays in chimpanzees. *Int J Primatol* 26(1):73–103.
24. Ghazanfar AA, Takahashi DY, Mathur N, Fitch WT (2012) Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Curr Biol* 22(13): 1176–1182.
25. Ferrari PF, Paukner A, Ionica C, Suomi SJ (2009) Reciprocal face-to-face communication between rhesus macaque mothers and their newborn infants. *Curr Biol* 19(20): 1768–1772.
26. Ferrari PF, et al. (2006) Neonatal imitation in rhesus macaques. *PLoS Biol* 4(9):e302.
27. De Marco A, Visalberghi E (2007) Facial displays in young tufted Capuchin monkeys (*Cebus apella*): Appearance, meaning, context and target. *Folia Primatol (Basel)* 78(2): 118–137.
28. Ghazanfar AA, Chandrasekaran C, Morrill RJ (2010) Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: Implications for the evolution of audiovisual speech. *Eur J Neurosci* 31(10):1807–1817.
29. Morrill RJ, Paukner A, Ferrari PF, Ghazanfar AA (2012) Monkey lipsmacking develops like the human speech rhythm. *Dev Sci* 15(4):557–568.
30. Chandrasekaran C, Ghazanfar AA (2009) Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. *J Neurophysiol* 101(2): 773–788.
31. Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J Neurosci* 28(17):4457–4469.
32. Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25(20): 5004–5012.
33. Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46(2):369–384.
34. Lakatos P, et al. (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol* 94(3):1904–1911.
35. Kayser C, Montemurro MA, Logothetis NK, Panzeri S (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron* 61(4):597–608.
36. Chandrasekaran C, Lemus L, Trubanova A, Gondon M, Ghazanfar AA (2011) Monkeys and humans share a common computation for face/voice integration. *PLoS Comput Biol* 7(9):e1002165.
37. Steckenfinger SA, Ghazanfar AA (2009) Monkey visual behavior falls into the uncanny valley. *Proc Natl Acad Sci USA* 106(43):18362–18366.
38. Lewkowicz DJ (1985) Developmental changes in infants' response to temporal frequency. *Dev Psychobiol* 21(2):858–865.
39. Zangenehpour S, Ghazanfar AA, Lewkowicz DJ, Zatorre RJ (2009) Heterochrony and cross-species intersensory matching by infant vervet monkeys. *PLoS ONE* 4(1):e4302.
40. Karmel BZ, Lester ML, McCarvill SL, Brown P, Hofmann MJ (1977) Correlation of infants' brain and behavior response to temporal changes in visual stimulation. *Psychophysiology* 14(2):134–142.
41. Ghazanfar AA, Logothetis NK (2003) Neuroperception: Facial expressions linked to monkey calls. *Nature* 423(6943):937–938.
42. Ghazanfar AA, et al. (2007) Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr Biol* 17(5):425–430.
43. Jordan KE, Brannon EM, Logothetis NK, Ghazanfar AA (2005) Monkeys match the number of voices they hear to the number of faces they see. *Curr Biol* 15(11): 1034–1038.
44. Maier JX, Neuhoff JG, Logothetis NK, Ghazanfar AA (2004) Multisensory integration of looming signals by rhesus monkeys. *Neuron* 43(2):177–181.
45. Giraud AL, et al. (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56(6):1127–1134.
46. Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Front Psychol* 3:1–12.
47. Kayser C, Logothetis NK, Panzeri S (2010) Visual enhancement of the information representation in auditory cortex. *Curr Biol* 20(1):19–24.
48. Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. *Science* 304(5679):1926–1929.
49. Karmel BZ, Maisel EB (1975) A neuronal activity model of infant visual attention. *Infant Perception: From Sensation to Cognition. Part I: Basic Visual Processes*, eds Cohen LB, Salapatek P (Academic Press, New York), pp 77–131.
50. Arcadi AC, Robert D, Boesch C (1998) Buttress drumming by wild chimpanzees: Temporal patterning. Phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates* 39(4):505–518.
51. Kalan AK, Rainey HJ (2009) Hand-clapping as a communicative gesture by wild female swamp gorillas. *Primates* 50(3):273–275.
52. Mehlman PT (1996) Branch shaking and related displays in wild Barbary macaques. *Evolution and Ecology of Macaque Societies*, eds Fa JE, Lindburg DG (Cambridge Univ. Press, Cambridge, UK), pp 503–526.
53. Remedios R, Logothetis NK, Kayser C (2009) Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proc Natl Acad Sci USA* 106(42):18010–18015.
54. Hiimeae KM, Palmer JB (2003) Tongue movements in feeding and speech. *Crit Rev Oral Biol Med* 14(6):413–429.
55. Hiimeae KM, et al. (2002) Hyoid and tongue surface movements in speaking and eating. *Arch Oral Biol* 47(1):11–27.
56. Moore CA, Smith A, Ringel RL (1988) Task-specific organization of activity in human jaw muscles. *J Speech Hear Res* 31(4):670–680.
57. Ostry DJ, Munhall KG (1994) Control of jaw orientation and position in mastication and speech. *J Neurophysiol* 71(4):1528–1545.
58. Matsuo K, Palmer JB (2010) Kinematic linkage of the tongue, jaw, and hyoid during eating and speech. *Arch Oral Biol* 55(4):325–331.
59. Locke JL (2008) Lipsmacking and babbling: Syllables, sociality, and survival. *The Syllable in Speech Production*, eds Davis BL, Zajdo K (Lawrence Erlbaum Associates, New York), pp 111–129.
60. Moore CA, Ruark JL (1996) Does speech emerge from earlier appearing oral motor behaviors? *J Speech Hear Res* 39(5):1034–1047.
61. Steeve RW (2010) Babbling and chewing: Jaw kinematics from 8 to 22 months. *J Phonetics* 38(3):445–458.
62. Steeve RW, Moore CA, Green JR, Reilly KJ, Ruark McMurtrey J (2008) Babbling, chewing, and sucking: Oromandibular coordination at 9 months. *J Speech Lang Hear Res* 51(6):1390–1404.