

# Cineradiography of Monkey Lip-Smacking Reveals Putative Precursors of Speech Dynamics

Asif A. Ghazanfar,<sup>1,2,3,\*</sup> Daniel Y. Takahashi,<sup>1,3</sup> Neil Mathur,<sup>1,2</sup> and W. Tecumseh Fitch<sup>4,\*</sup>

<sup>1</sup>Neuroscience Institute

<sup>2</sup>Departments of Ecology and Evolutionary Biology

<sup>3</sup>Departments of Psychology

Princeton University, Princeton, NJ 08540, USA

<sup>4</sup>Department of Cognitive Biology, Faculty of Life Sciences, University of Vienna, 14 Althanstrasse, A-1090 Vienna, Austria

## Summary

A key feature of speech is its stereotypical 5 Hz rhythm [1, 2]. One theory posits that this rhythm evolved through the modification of rhythmic facial movements in ancestral primates [3, 4]. If the hypothesis has any validity, then a comparative approach may shed some light. We tested this idea by using cineradiography (X-ray movies) to characterize and quantify the internal dynamics of the macaque monkey vocal tract during lip-smacking (a rhythmic facial expression) versus chewing. Previous human studies showed that speech movements are faster than chewing movements, and the functional coordination between vocal tract structures is different between the two behaviors [5–9]. If rhythmic speech evolved through a rhythmic ancestral facial movement, then one hypothesis is that monkey lip-smacking versus chewing should also exhibit these differences. We found that the lips, tongue, and hyoid move with a speech-like 5 Hz rhythm during lip-smacking, but not during chewing. Most importantly, the functional coordination between these structures was distinct for each behavior. These data provide empirical support for the idea that the human speech rhythm evolved from the rhythmic facial expressions of ancestral primates.

## Results

Speech acoustics typically exhibit a 3–8 Hz rhythm, related to the rate of syllable production. This rhythm is critical for speech perception [10, 11], possibly because it exploits brain rhythms in the same frequency range [12–14]. Understanding how speech evolved requires investigating its rhythmic origins. Because most traits involved in speech—the vocal production apparatus and the brain—do not fossilize [15, 16], we are left with only the comparative method for investigating the evolution of speech. By comparing the behavior and biology of extant primates with humans, we can deduce the behavioral capacities of extinct common ancestors.

To test the hypothesis that the rhythmic structure of human speech evolved from the rhythmic facial expressions of ancestral primates, we used cineradiography to investigate whether vocal tract movements during macaque monkey lip-smacking are similar to vocal tract movements during human speech and different from those produced during chewing. Lip-smacking

is an affiliative facial expression observed in many Old World primates [17–19]. It is characterized by bouts of regular cycles of vertical jaw movement. Importantly, lip-smacking is almost always directed at another individual during face-to-face interactions and sometimes appears to involve turn-taking, much like speech [18, 20].

The vocal tract structures used for speech production overlap completely with those involved in ingestion, and cineradiography offers an unparalleled window into their internal movements. The method can be applied across species [21, 22]. As in previous cineradiographic studies of human speech versus chewing [5, 6, 9], we focused on the tongue, lips, and hyoid bone (an integral laryngeal structure) (Figures 1A and 1B). To collect the relevant image sequences during cineradiography, we relied upon spontaneous bouts of lip-smacking during interactions with experimenters or with other monkeys. The analyses below are based on data collected from three monkeys: ten lip-smacking bouts (six from monkey 1 [M1], four from M3) and ten chewing bouts (four from M1, two from M2, and four from M3).

## Spatial Displacement of Vocal Tract Elements

We mapped the spatial displacement of the lips, tongue, and hyoid during the production of lip-smacking and chewing movements. To compare their spatial distribution of positions across different subjects with varying head sizes, we normalized the data set such that the distance between the upper hyoid marker and lower tongue marker was always one (Figure 1B). The coordinates were labeled (0,0) for the upper hyoid marker and (1,0) for the lower tongue marker. Figures 2A and 2B show the distribution of displacements of the three structures during lip-smacking and chewing. For the most part, these distributions look the same for the two behaviors, but the excursions that the structures take are greater for chewing than for lip-smacking. This is more apparent in the centroids (with 95% confidence intervals) shown in Figures 2C and 2D. Overall, however, the displacement of the tongue, lips, and hyoid are similar between chewing and lip-smacking. These data are consistent with what is seen for chewing and speech in humans: greater excursions of the tongue, jaw, and hyoid during chewing than during speech [6]. However, during speech but not lip-smacking, the hyoid occupies a space that is distinct from that seen during chewing [6].

## Temporal Structure of Vocal Tract Movements

Human speech has a rhythm that is ~5 Hz [1, 2] and human chewing movements are highly stereotyped with a slower frequency of ~2.5 Hz [6, 23]. We know from previous work that adult macaque monkeys move their lips during lip-smacking at a rate of ~5 Hz [24, 25], but it remained unknown whether or not other internal vocal tract elements (tongue and hyoid) do so as well. Figure 3A shows the time series of two lip-smacking examples from two different monkeys. All three vocal tract elements seem to oscillate at a frequency between 4 and 5 Hz. A spectral analysis on the entire sample (n = 10 lip-smacking bouts) reveals that the average frequency of tongue oscillation was 4.56 Hz, the average lip oscillation was 4.50 Hz, and the average hyoid oscillation was 3.80 Hz (Figure 3C).

\*Correspondence: asifg@princeton.edu (A.A.G.), tecumseh.fitch@univie.ac.at (W.T.F.)

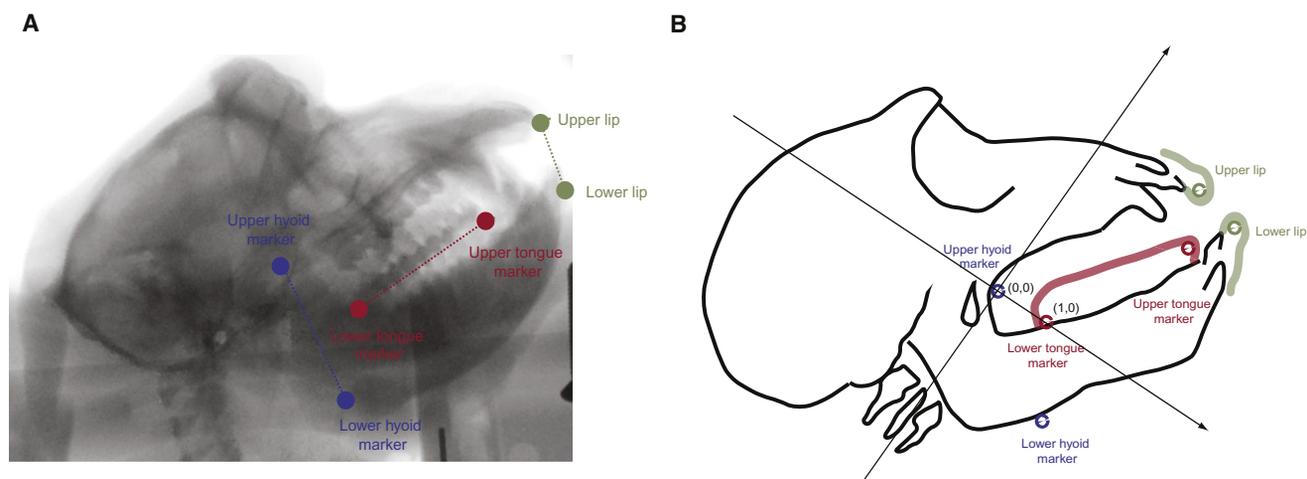


Figure 1. The Anatomy of the Macaque Monkey Vocal Tract as Imaged with Cineradiography  
(A) Shown are identifying marks for the key vocal tract structures, including the lips, tongue, and hyoid.  
(B) Shown is a schematic of the vocal tract elements and the reference points for measurements.

The chewing rhythm was slower (Figure 3B). The average frequency of tongue oscillation was 2.13 Hz, the average lip oscillation was 2.29 Hz, and the average hyoid oscillation was 1.88 Hz (Figure 3D). These results almost exactly parallel what is known for human speech and chewing. For each structure in macaque monkeys, the differences in the rhythmic frequency during lip-smacking versus chewing were statistically significant (lip,  $p < 0.001$ ; tongue,  $p < 0.001$ ; hyoid,  $p < 0.001$ ).

### Coordination of Vocal Tract Elements

Thus far, our data show similarities between the spatial and temporal characteristics of vocal tract motion during monkey lip-smacking and human speech, and monkey chewing and human chewing. Our next analysis focused on the temporal coordination of the three vocal tract elements during lip-smacking versus chewing. Previous cineradiography studies showed that, for chewing, there is a robust coupling of the jaw, tongue, and hyoid in humans, monkeys, and other animals [6, 9, 26, 27]. In contrast, for human speech production, there is less coupling between the hyoid, jaw, and tongue movements [6, 9]. To quantify the strength of coupling between vocal tract elements during lip-smacking versus chewing in monkeys, we analyzed our time-series using partial directed coherence (PDC). PDC is an established method to infer the information flow, or the causal ordering, between time series [28]. It measures how much of the information in one time series can be explained by another time-series.

We generated PDC plots of each of the three tract components against each other for lip-smacking and chewing (Figures 4A and 4B). The higher the PDC value, the more likely it is that the source component predicts the receiver component at a given frequency. The black line in each subplot shows the frequency-dependent PDC value, and the brown-shaded regions denote confidence intervals. Figure 4A shows that for lip-smacking, with the lips as the source and the tongue as the receiver, a significant peak occurred at 4.10 Hz (green shaded region;  $\text{PDC} = 0.72$ ,  $p \text{ value} < 0.01/6$ ). Similarly, with the tongue as the source and the lips as the receiver, a significant peak occurred at 3.28 Hz (red shaded region;  $\text{PDC} = 0.45$ ,  $p \text{ value} < 0.01/6$ ). Figure 4B shows that during chewing, with

the tongue as the source and the lips as the receiver, a significant peak occurred at 2.81 Hz (red shaded region;  $\text{PDC} = 0.32$ ,  $p \text{ value} < 0.01/6$ ), and with the lips as the source and the hyoid as the receiver, a significant peak occurred at 2.93 Hz (green shaded region;  $\text{PDC} = 0.37$ ,  $p \text{ value} < 0.01/6$ ). In contrast, for lip-smacking, “lips to hyoid,” “tongue to hyoid,” “hyoid to lips,” and “hyoid to tongue” are all lacking in significant relationship (Figure 4A), and during chewing movements, “lips to tongue,” “tongue to hyoid,” “hyoid to lips,” and “hyoid to tongue” are similarly not significant (Figure 4B).

Overall, the pattern of coupling between vocal tract structures is different during lip-smacking and chewing in a way comparable to the differences between speech and chewing. Although all three vocal tract structures are moving, during lip-smacking there is strong bidirectional coupling between the tongue and the lips, but no coupling of either structure with the hyoid (Figure 4A, center panel). This is consistent with what is reported for speech production in humans: although the hyoid moves continuously during speech, there is no coupling of the hyoid with jaw movements [6] and generally less coordination among all three structures when compared to chewing [9]. In both species, this lack of coupling with the hyoid is in contrast with chewing movements during which the three structures are functionally coupled and have a feed-forward relationship with each other (Figure 4B, center panel) [6, 9].

### Discussion

We found that while the spatial displacement of the vocal tract elements (lips, tongue, and hyoid) occupied a similar active space during lip-smacking and chewing, when temporal interactions were analyzed, the two rhythmical movements clearly differed in the same manner that human speech and chewing differ. Specifically, we show that the lips, tongue, and hyoid all move at  $\sim 4\text{--}5$  Hz during lip-smacking, which is consistent with the universal rhythm of speech [1, 2, 29] and distinct from the rhythm of chewing movements in both monkeys and humans ( $\sim 2.5$  Hz; [6, 23, 25, 30]). Importantly, we also show that the coordination of the vocal tract elements during lip-smacking and speech production is similar: in both cases,

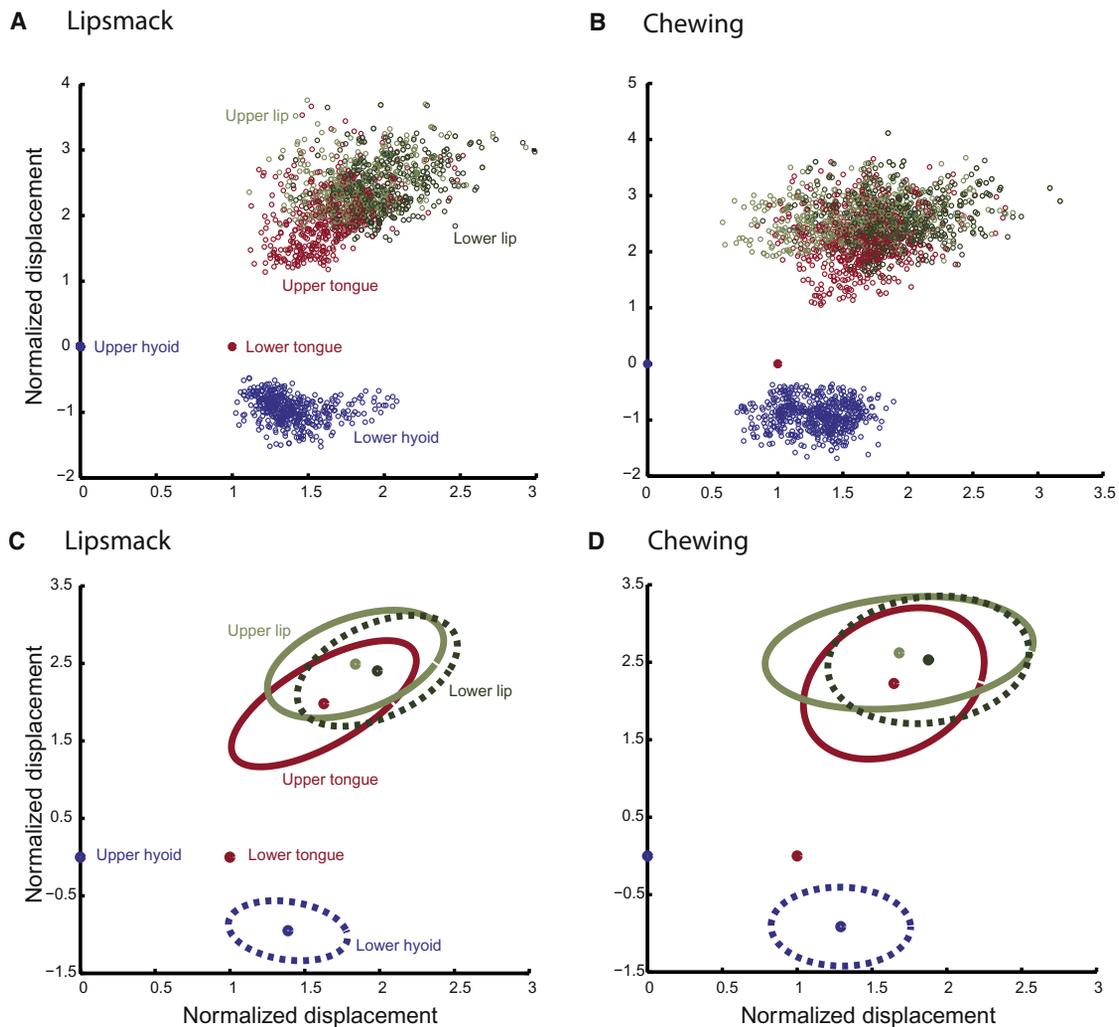


Figure 2. Relative Spatial Distribution of the Lips, Tongue and Hyoid

(A) Shown is the scatterplot for the relative spatial distribution of the upper lip (light green), lower lip (dark green), upper and lower tongues (red), upper and lower hyoid (blue), during lip-smacking, for all frames, all sessions, and all monkeys. Observe that the scale is normalized such that the upper hyoid is always at the origin (coordinate [0,0]) and the lower tongue is always at position (1,0).

(B) Shown is the scatterplot for the same structures as (A) but during chewing.

(C and D) Shown are the centroids and 95% confidence intervals for the spatial distributions of different structures shown in (A) and (B), respectively.

the hyoid is not coupled to the lips and tongue [6, 9]. This is in contrast to chewing (for both species) where all three are linked to each other's movements. Overall, our data show striking similarities between vocal tract movements during monkey lip-smacking and human speech and lend strong empirical support for the idea that the dynamic control of human speech articulators evolved from the rhythmic facial expressions of a common ancestor to both humans and macaques [3, 4].

Our data show that the similarities between monkey lip-smacking and human speech go beyond the visible orofacial (mandibular) movements and include the dynamics of the tongue and hyoid. In light of this, MacNeilage's hypothesis [3, 4] in which the role of mandibular oscillations is central must be modified. Any theory of speech evolution invoking ancestral rhythmic facial expression must take into account the complex biomechanical relationship between the jaw, lips, hyoid, and tongue, as well as their distinct neural pathways.

#### Differences between Lip-Smacking and Speech Production

Our data show that two core features of speech production—its rhythmical structure and temporal coordination of vocal elements—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 component (though a quiet consonant-like bilabial plosive or /p/ sound is produced when the lips smack together). 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 [16, 31, 32]. 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

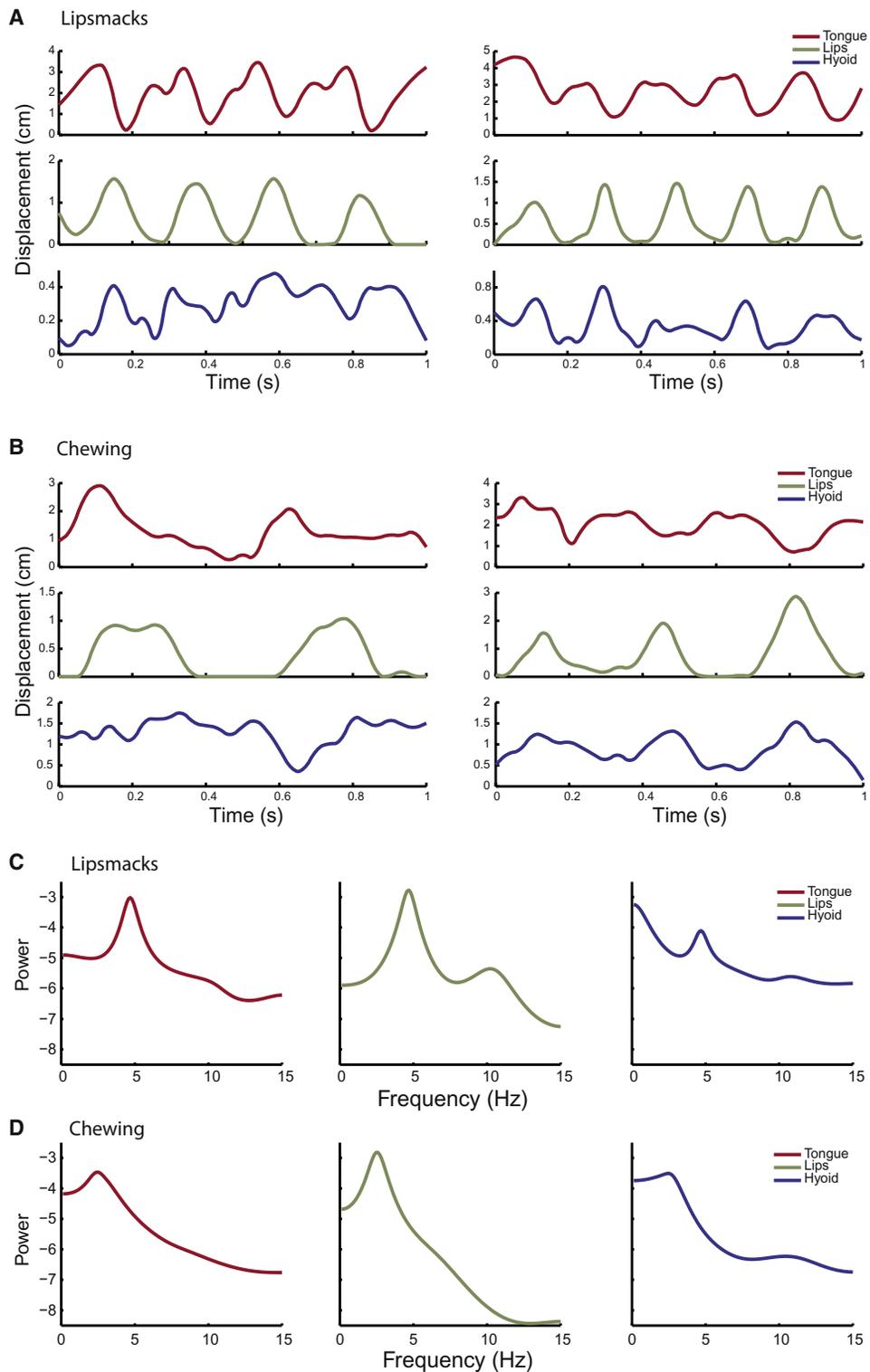
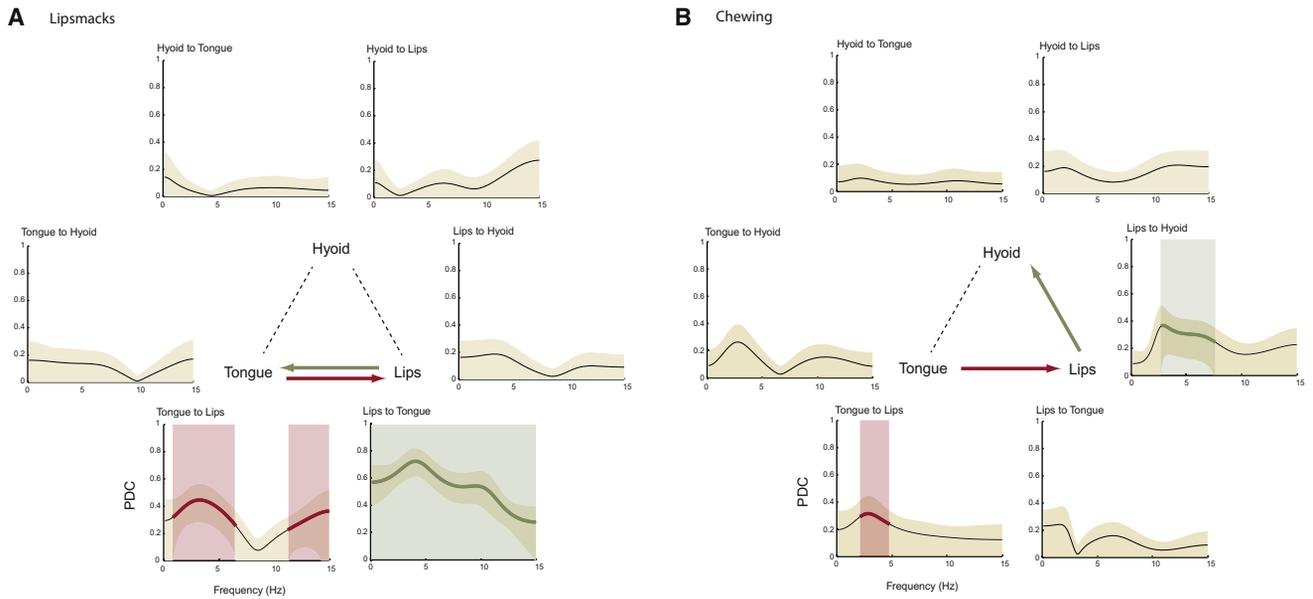


Figure 3. Rhythmic Structure of Orofacial Movements

Time-displacement plot of the tongue, interlip distance, and hyoid for two examples of lip-smacking from two different monkeys (A) and chewing from two different monkeys (B). The x axis depicts the time in s and the y axis the displacement in cm. Power spectra of the time-varying displacement of the tongue, interlip distance, and hyoid during lip-smacking (C) and chewing (D) are shown. x axis depicts the frequency in hertz, and the y axis depicts the value of the logarithm of power spectrum in normalized units. The upper plots show the power spectrum of the displacement of tongue (red), interlip distance (green), and hyoid (blue). The lower plots show the power spectrum of the same structures during the chewing.



**Figure 4. Temporal Coordination of Vocal Tract Elements**

(A) PDC plots between tongue, interlip distance, and hyoid for lip-smacking. x axis depicts the frequency in hertz and the y axis the PDC value in normalized units (between 0 and 1). The solid black line in each plot indicates the PDC value for the respective pair of structures. The brown area corresponds to the 95% confidence intervals. The significant PDC values with p values < 0.01/6 are shown in red (tongue to lips) and in green (lips to tongue). The arrows in the center depict the direction of significant influence from each structure onto to the other two as measured by the PDC analysis. (B) PDC plots between tongue, lips, and hyoid for the chewing. The conventions follow (A).

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. Our data only address the evolution of vocal tract movements (the filter component) involved in speech production.

Other differences between lip-smacking and speech are the range of hyoid movements and the coupling of the lips with the tongue. For the latter, the coupling of the lips and tongue during lip-smacking (Figure 4) is unlikely to be the case for human speech where their independence allows for the production of a wide range of sounds (though this hasn't been tested explicitly). For the former, our data show that the hyoid occupies the same active space during lip-smacking and chewing (Figure 2). In contrast, cineradiography studies of human speech versus chewing show a dichotomy in hyoid movement patterns [6]. These movement range differences of the hyoid in humans versus macaques could be due to 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 shifts [33]. This increases the range of tongue movements, and possibly hyoid movements, relative to nonhuman primates. Movements of either or both effectors could influence the active space of the hyoid.

#### Neural Control of Lip-Smacking and the Speech Rhythm

The vocal tract movements shared by chewing, lip-smacking, and speech all require the coordination of muscles controlling the jaw, hyoid, tongue, and respiration. Homologous central pattern generators in the brainstem likely produce their foundational rhythms and are modulated by feedback from peripheral sensory receptors. Beyond sensory feedback, one

additional source influencing how differences between orofacial movements may arise is the neocortex. Although chewing movements may be largely independent of cortical control [34], lip-smacking and speech production are both modulated by the neocortex. Thus, one hypothesis is that the differences between the frequency of lip-smacking and speech versus chewing reflect the influence of neocortical circuits upon brainstem central pattern generators. In addition to the superior temporal sulcus [24] and the premotor cortex [35], one other important neocortical node likely to be involved is the insula. The human insula is involved in multiple processes related to communication, including speech production [36, 37]. Consistent with an evolutionary link between lip-smacking and speech, the insula also plays a role in monkey lip-smacking [38].

#### Experimental Procedures

##### Subjects

Subjects were three adult male long-tailed macaque monkeys, *Macaca fascicularis*. Monkeys were seated in restraint chairs. All experiments were performed in compliance with the Institutional Animal Care and Use Committee guidelines for the care and use of laboratory animals.

##### Cineradiography

Cineradiography was performed using a Philips Pulsera C-arm with a 12-inch image intensifier; this is a mobile X-ray machine that captures videos at 30 frames per s. Monkeys were seated in restraint chairs during imaging but were not head-fixed and had a complete range of head motion essential to capturing natural orofacial gestures. For reasons of radiation safety, the Philips system automatically shut off its generation of X-rays after 10 s. Thus, the temporal limit of our longest recordings was 10 s. To elicit lip-smacking, the experimenter would lip-smack toward the subject and/or another monkey in a restraint chair was brought into the room and placed in front of the subject. Lip-smacking is an unambiguous affiliative signal [17, 18]. Chewing movements were elicited by feeding the subject grapes or pieces of banana. At the beginning of an imaging session and

during pauses within the session, monkeys were given a 2% barium sulfate suspension to outline the vocal tract contour.

For each imaging session, a calibration grid was briefly presented near the monkey's head to provide accurate measurements during image processing.

#### Image Processing and Tracing

Each video was captured using a Canopus ADVC110 audio/video digitizer and later decomposed into a series of DICOM PNG frames. Only those videos with a clear sagittal view of the monkey's head and throat were used. Midsagittal alignment was important because measures of the effectors (lips, tongue, and hyoid) depended on consistent positioning. Any roll or yaw of the head would affect the measurement of the hyoid-to-palate distance and the measurement of tongue length, respectively. To control for these potential misalignments, we made sure that the canine teeth and the ocular ridges were aligned in each video used for analyses. This important requirement severely reduced the number of useable videos. From 260 recordings, we narrowed down our selection of videos to 10 lip-smacking and 10 chewing bouts from three monkeys. These were selected for consistent midsagittal alignment and edited to include only the relevant orofacial movements. Bouts of mouth movements were always >1.5 s in duration.

Image frames were subsequently imported into Adobe Photoshop CS4. Different color-coded tracing layers for different anatomical structures were created for each image using a Wacom Cintiq 12WX tablet. To trace the individual vocal tract components, we first created a layer in which the contrast was toggled to its maximum, and then separately created a layer in which the brightness was toggled to its minimum. The former clarified the positions of the hyoid, whereas the latter allowed for visualization of the lips and tongue. Tracing the tongue and lips was straightforward. For the hyoid tracings, the curve of the hyoid was followed down by the thyroid cartilage and then perpendicularly across the throat to the interarytenoid muscle next to the visible cricoid cartilage. The posterior line of the vertical portion of the supralaryngeal vocal tract was followed upwards from the interarytenoid muscle to the tip of the velum at the end of the soft palate. On the anterior side, the vocal tract followed the thyroid cartilage up past the hyoid bulla and around the curve of the epiglottis to the base of the tongue.

#### Measuring Spatial and Temporal Dynamics

Once the tracings were complete, we reconstructed the video using the traced image frames. A custom Matlab script was used to measure the distances between different vocal tract components as a function of time. With each anatomically labeled image, we sequentially specified points for the anterior base of the tongue to its tip (tongue length), the soft palate down to the vocal folds at the top of the hyoid, the interlip distance, and the length of the velum from the soft palate to its tip. We included a calibration process to convert the measured pixel distances into centimeters.

#### Partial Directed Coherence

We used partial directed coherence (PDC) to elucidate the temporal coordination between the selected vocal tract components [28]. The custom Matlab scripts for computing the PDC analyses are available here: <http://www.lcs.poli.usp.br/~baccala/pdc/software/asymppdc.zip>. Time series from different monkeys were first normalized to avoid the confounding variable of vocal tract size differences between monkeys. Bouts of mouth movements (either lip-smacking or chewing) were then concatenated together. This was done for statistical robustness. Because lip-smacking and chewing bouts were of short and differing durations and sampled at the low rate of 30 Hz, estimation and inference for trivariate analyses (lip, tongue, and hyoid movements) would be difficult if individual bouts were used. Moreover, averaging the PDC values across individual movement bouts is problematic because the data segments have different lengths and the weight of PDC values would increase with segment length—a confound. The concatenation takes care of these statistical problems and is a standard practice in time-series analyses. The concatenated lip-smacking time series was 15.9 s in duration; the chewing time series was 18.7 s.

We further investigated the PDC data by finding the optimal maximum order of the autoregressive model to confirm robustness. To correct for multiple comparisons, we lowered the threshold used to test for significance to  $\alpha = 0.01/6$  because there were six directional relationships to examine between the three vocal tract components. Rather than have an additive threshold, we ensured that the significance throughout remained at an actual 1% threshold.

#### Acknowledgments

This work was supported by the National Institutes of Health (NINDS) R01NS054898 (A.A.G.), the Pew Latin American Fellowship (D.Y.T.), the Quantitative and Computational Neuroscience Training grant NIH R90 DA 023419-01 (N.M.), and the European Research Council Advanced Grant SOMACCA: Syntax of Mind: A Comparative Computational Approach (W.T.F.).

Received: January 1, 2012

Revised: March 14, 2012

Accepted: April 17, 2012

Published online: May 31, 2012

#### References

1. Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., and Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5, e1000436.
2. Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *J. Phonetics* 31, 465–485.
3. MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21, 499–511, discussion 511–546.
4. MacNeilage, P.F. (2008). *The Origin of Speech* (Oxford, UK: Oxford University Press).
5. Hiimae, K.M., and Palmer, J.B. (2003). Tongue movements in feeding and speech. *Crit. Rev. Oral Biol. Med.* 14, 413–429.
6. Hiimae, K.M., Palmer, J.B., Medicis, S.W., Hegener, J., Jackson, B.S., and Lieberman, D.E. (2002). Hyoid and tongue surface movements in speaking and eating. *Arch. Oral Biol.* 47, 11–27.
7. Moore, C.A., Smith, A., and Ringel, R.L. (1988). Task-specific organization of activity in human jaw muscles. *J. Speech Hear. Res.* 31, 670–680.
8. Ostry, D.J., and Munhall, K.G. (1994). Control of jaw orientation and position in mastication and speech. *J. Neurophysiol.* 71, 1528–1545.
9. Matsuo, K., and Palmer, J.B. (2010). Kinematic linkage of the tongue, jaw, and hyoid during eating and speech. *Arch. Oral Biol.* 55, 325–331.
10. Shannon, R.V., Zeng, F.-G., Kamath, V., Wyganski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304.
11. Smith, Z.M., Delgutte, B., and Oxenham, A.J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416, 87–90.
12. Giraud, A.L., Kleinschmidt, A., Poeppel, D., Lund, T.E., Frackowiak, R.S.J., and Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56, 1127–1134.
13. Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* 41, 245–255.
14. Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci. (Regul. Ed.)* 12, 106–113.
15. Fitch, W.T. (2000). The evolution of speech: a comparative review. *Trends Cogn. Sci. (Regul. Ed.)* 4, 258–267.
16. Ghazanfar, A.A., and Rendall, D. (2008). Evolution of human vocal production. *Curr. Biol.* 18, R457–R460.
17. Redican, W.K. (1975). Facial expressions in nonhuman primates. In *Primate Behavior: Developments in Field and Laboratory Research*, L.A. Rosenblum, ed. (New York: Academic Press), pp. 103–194.
18. Van Hooff, J.A.R.A.M. (1962). Facial expressions of higher primates. *Symposium of the Zoological Society* 8, 97–125.
19. Parr, L.A., Cohen, M., and de Waal, F. (2005). Influence of social context on the use of blended and graded facial displays in chimpanzees. *Int. J. Primatol.* 26, 73–103.
20. Ferrari, P.F., Paukner, A., Ionica, C., and Suomi, S.J. (2009). Reciprocal face-to-face communication between rhesus macaque mothers and their newborn infants. *Curr. Biol.* 19, 1768–1772.
21. Fitch, W.T. (2000). The phonetic potential of nonhuman vocal tracts: comparative cineradiographic observations of vocalizing animals. *Phonetica* 57, 205–218.
22. Riede, T., Suthers, R.A., Fletcher, N.H., and Blevins, W.E. (2006). Songbirds tune their vocal tract to the fundamental frequency of their song. *Proc. Natl. Acad. Sci. USA* 103, 5543–5548.

23. Green, J.R., Moore, C.A., Ruark, J.L., Rodda, P.R., Morvée, W.T., and VanWitzenburg, M.J. (1997). Development of chewing in children from 12 to 48 months: longitudinal study of EMG patterns. *J. Neurophysiol.* **77**, 2704–2716.
24. Ghazanfar, A.A., Chandrasekaran, C., and Morrill, R.J. (2010). Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: implications for the evolution of audiovisual speech. *Eur. J. Neurosci.* **31**, 1807–1817.
25. Morrill, R.J., Paukner, A., Ferrari, P.F., and Ghazanfar, A.A. (2012). Monkey lip-smacking develops like the human speech rhythm. *Dev. Sci.* Published online April 18, 2012. 10.1111/j.1467-7687.2012.01149.x.
26. German, R.Z., and Franks, H.A. (1991). Timing in the movement of jaws, tongue, and hyoid during feeding in the hyrax, *Procavia syriacus*. *J. Exp. Zool.* **257**, 34–42.
27. Hiiemae, K.M., Hayenga, S.M., and Reese, A. (1995). Patterns of tongue and jaw movement in a cinefluorographic study of feeding in the macaque. *Arch. Oral Biol.* **40**, 229–246.
28. Takahashi, D.Y., Baccala, L.A., and Sameshima, K. (2007). Connectivity inference between neural structures via partial directed coherence. *J. Appl. Stat.* **34**, 1259–1273.
29. Dolata, J.K., Davis, B.L., and Macneilage, P.F. (2008). Characteristics of the rhythmic organization of vocal babbling: implications for an amodal linguistic rhythm. *Infant Behav. Dev.* **31**, 422–431.
30. Kiliaridis, S., Karlsson, S., and Kjellberg, H. (1991). Characteristics of masticatory mandibular movements and velocity in growing individuals and young adults. *J. Dent. Res.* **70**, 1367–1370.
31. Fitch, W.T., and Hauser, M.D. (1995). Vocal production in nonhuman-primates - acoustics, physiology, and functional constraints on honest advertisement. *Am. J. Primatol.* **37**, 191–219.
32. Fant, G. (1970). *Acoustic Theory of Speech Production*, Second Edition (Paris: Mouton).
33. Lieberman, D.E., McCarthy, R.C., Hiiemae, K.M., and Palmer, J.B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Arch. Oral Biol.* **46**, 117–128.
34. Lund, J.P., and Kolta, A. (2006). Brainstem circuits that control mastication: do they have anything to say during speech? *J. Commun. Disord.* **39**, 381–390.
35. Ferrari, P.F., Gallese, V., Rizzolatti, G., and Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur. J. Neurosci.* **17**, 1703–1714.
36. Ackermann, H., and Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain Lang.* **89**, 320–328.
37. Dronkers, N.F. (1996). A new brain region for coordinating speech articulation. *Nature* **384**, 159–161.
38. Caruana, F., Jezzini, A., Sbriscia-Fioretta, B., Rizzolatti, G., and Gallese, V. (2011). Emotional and social behaviors elicited by electrical stimulation of the insula in the macaque monkey. *Curr. Biol.* **21**, 195–199.