Monkey Lip-smacking Develops Like the Human Speech Rhythm

Ryan J. Morrill[1,2], **Annika Paukner**[3], **Pier F. Ferrari**[4], and **Asif A. Ghazanfar**[1,2,5,*]

[1]Neuroscience Institute, Princeton University, Princeton NJ 08540, USA

[2]Department of Ecology & Evolutionary Biology, Princeton University, Princeton NJ 08540, USA

[3]Laboratory of Comparative Ethology, Eunice Kennedy Shriver National Institute of Child Health and Human, Development, National Institutes of Health, Bethesda MD, 20892-7971, USA

[4]Dipartimento di Biologia Evolutiva e Funzionale and Dipartimento di Neuroscienze, Via Usberti 11A, Università di Parma, 43100 Parma, Italy

[5]Department of Psychology, Princeton University, Princeton NJ 08540, USA

Abstract

Across all languages studied to date, audiovisual speech exhibits a consistent rhythmic structure. This rhythm is critical to speech perception. Some have suggested that the speech rhythm evolved *de novo* in humans. An alternative account—the one we explored here—is that the rhythm of speech evolved through the modification of rhythmic facial expressions. We tested this idea by investigating the structure and development of macaque monkey lipsmacks and found that their developmental trajectory is strikingly similar to the one that leads from human infant babbling to adult speech. Specifically, we show that: 1) younger monkeys produce slower, more variable mouth movements and as they get older, these movements become faster and less variable; and 2) this developmental pattern does *not* occur for another cyclical mouth movement—chewing. These patterns parallel human developmental patterns for speech and chewing. They suggest that, in both species, the two types of rhythmic mouth movements use different underlying neural circuits that develop in different ways. Ultimately, both lipsmacking and speech converge on a ~5 Hz rhythm that represents the frequency that characterizes the speech rhythm of human adults. We conclude that monkey lipsmacking and human speech share a homologous developmental mechanism, lending strong empirical support for the idea that the human speech rhythm evolved from the rhythmic facial expressions of our primate ancestors.

Keywords

Introduction

Determining how human speech evolved is difficult primarily because most traits thought to give rise to speech — the vocal production apparatus and the brain — do not fossilize. We are left with only one reliable way of investigating the mechanisms underlying the evolution of speech: the comparative method. By comparing the behavior and biology of extant primates with humans, we can deduce the behavioral capacities of extinct common ancestors, allowing identification of homologies and providing clues as to the adaptive functions of these behaviors. However, comparative studies must also recognize that

*To whom correspondence should be addressed: asifg@princeton.edu.

species-typical behaviors are not only the product of phylogenetic processes but ontogenetic ones as well (Gottlieb 1992), and thus understanding the origins of species-typical behaviors requires understanding the relationship between these two processes. This integrative approach can help determine whether homologies reflect the operation of the same or different underlying mechanisms (Deacon 1990; Finlay et al. 2001; Schneirla 1949). Thus, a deeper understanding of how speech evolved requires comparative studies that incorporate the developmental trajectories of putative speech-related behaviors.

Across all languages studied to date, audiovisual speech exhibits a ~5 Hz rhythmic structure (Chandrasekaran et al. 2009; Crystal & House 1982; Dolata et al. 2008; Greenberg et al. 2003; Malecot et al. 1972). This ~5 Hz rhythm is critical to speech perception. Disrupting the auditory component of this rhythm significantly reduces intelligibility (Drullman et al. 1994; Saberi & Perrott 1999; Shannon et al. 1995; Smith et al. 2002), as does disrupting the visual (visible mouth movements) component (Campbell 2008; Kim & Davis 2004; Vitkovitch & Barber 1994; Vitkovitch & Barber 1996). The speech rhythm is also closely related to on-going brain rhythms which it may exploit through a common sampling rate or by entrainment (Giraud et al. 2007; Luo et al. 2010; Luo & Poeppel 2007; Poeppel 2003; Schroeder et al. 2008). These brain rhythms are common to all mammals (Buzsaki & Draguhn 2004). While some have suggested that the rhythm of speech evolved *de novo* in humans (Pinker & Bloom 1990), one influential theory posits that it evolved through the modification of rhythmic facial movements in ancestral primates (MacNeilage 1998; MacNeilage 2008).

Rhythmic facial movements are extremely common as visual communicative gestures in primates. The lipsmack, for example, is an affiliative signal observed in many genera of Old World primates (Hinde & Rowell 1962; Redican 1975; Van Hooff 1962), including chimpanzees (known as "teeth-clacks"; (Goodall 1968; Parr et al. 2005)). In macaque monkeys, it is characterized by regular cycles of vertical jaw movement, often involving a parting of the lips, but sometimes occurring with closed, puckered lips. Importantly, as a communication signal, the lipsmack is typically directed at another individual during face-to-face interactions (Ferrari et al. 2009; Van Hooff 1962). Thus, during the course of speech evolution, as the theory goes, these rhythmic facial expressions were coupled to vocalizations to produce the audiovisual components of babbling-like (consonant-vowel-like) expressions (MacNeilage 1998; MacNeilage 2008). Tests of such evolutionary hypotheses are difficult. Yet, if the idea that rhythmic speech evolved through the rhythmic facial expressions of ancestral primates has any validity, then there are at least three predictions that can be tested using the comparative approach. The first is that, like speech, rhythmic facial expressions of extant primates should occur with a ~5 Hz frequency (Ghazanfar et al. 2010). The second, stronger prediction is that, if the underlying mechanisms that produce the rhythm in monkey lipsmacks and human speech are homologous, then their developmental trajectories should be similar (Gottlieb 1992; Schneirla 1949). Finally, this common trajectory should be distinct from the developmental trajectory of other rhythmic mouth movements.

In humans, the earliest form of rhythmic vocal behavior occurs some time after six-months of age, when vocal babbling abruptly emerges (Locke 1993; Preuschoff et al. 2008; Smith & Zelaznik 2004). 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 (Davis & MacNeilage 1995; Lindblom et al. 1996; Oller 2000). 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 at least

two aspects to these changes: frequency and variability. In adults, the speech rhythm is ~5 Hz (Chandrasekaran et al. 2009; Crystal & House 1982; Dolata et al. 2008; Greenberg et al. 2003; Malecot et al. 1972), while in infant babbling, the rhythm is considerably slower. Between 2 and 12 months of age, infants produce speech-like sounds at a slower rate of roughly 2.8 to 3.4 Hz (Dolata et al. 2008; Levitt & Wang 1991; Lynch et al. 1995; Nathani et al. 2003). 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 post-pubescence (Smith & Zelaznik 2004). Importantly, this developmental trajectory from babbling to speech is distinct from that of another cyclical mouth movement, that of *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 et al. 1991).

If lipsmacking is indeed a plausible substrate for the evolution of speech, then it should emerge ontogenetically like human speech. Here, we tested this hypothesis by measuring the rhythmic frequency and variability of lipsmacking in macaque monkeys across neonatal, juvenile and adult age groups. Though great apes are phylogenetically closer to humans, we used macaques as our model species because it is very difficult (if not impossible) to acquire developmental data from great apes. There were many possible outcomes to our study. First, given the differences in the size of the facial structures between macaques and humans, it is possible that lipsmacks and speech rhythms don't converge on the same ~5 Hz rhythm. Second, because of the precocial neocortical development of macaque monkeys relative to humans (Gibson 1991; Malkova et al. 2006), the lipsmack rhythm could remain stable from birth onwards and show no changes in frequency and/or no changes in variability (Thelen 1981). Finally, whatever changes in lipsmack structure may occur, the developmental trajectory may be similar to that of chewing, another rhythmic mouth movement. That is, it may be that all complex motor acts undergo the same changes in frequency and variablity. Our data show that, like human speech development, monkey lipsmacking transitions from low frequency and high variability state to higher frequency and lower variability over the course of development. Both communication signals converge on ~ 5 Hz. Moreover, as in human speech, the developmental trajectory of lipsmacking is different from the trajectory for chewing-related orofacial movements. Our data suggest that monkey lipsmacking and human speech share a homologous developmental mechanism and provide empirical support for the idea that human speech evolved from the rhythmic facial expressions of our primate ancestors (MacNeilage 1998; MacNeilage 2008).

Methods

All testing of nonhuman primates was conducted in accordance with regulations governing the care and use of laboratory animals and had prior approval from the Institutional Animal Care and Use Committee of the National Institute of Child Health and Human Development and the University of Puerto Rico School of Medicine who operate on behalf of the Caribbean Primate Research Center.

Classification of lipsmacks

The lipsmack is an unambiguous affiliative signal consisting of rapid vertical displacement of the lower jaw and a puckering of the lips (Hinde & Rowell 1962; Redican 1975; Van Hooff 1962)(Figure 1). The upper and lower teeth do not come together during the lipsmack, as they do for teeth-grind expressions. Occasionally, the tongue protrudes slightly as the jaw opens and the lips part. Importantly, the lipsmack display is typical produced during face-to-face interactions and follows eye-contact (Ferrari et al. 2009; Van Hooff 1962). In all our

data collection, lipsmacks from monkeys were elicited through interactions with human experimenters mimicking lipsmacking behavior.

Neonatal lipsmack data collection

Neonatal rhythmic facial gestures were recorded from captive infant rhesus macaques between 3 and 8 days old (n=15), and in the context of a project investigating neonatal imitation (Figure 1A). All infants were nursery reared using procedures according to an established protocol (Ruppenthal et al. 1976). Infants were tested ~30–90 min after feedings in an experimental room in which the experimenter was seated on a chair and held the infant while it was grasping a surrogate (doll) mother or pieces of fleece fabric. This testing is routinely performed as part of a project aimed to assess neonatal imitative skills (Ferrari et al. 2006). Infants were tested three times a day for up to four days when 1–2 days old, 3–4 days old, 5–6 days old, and 7–8 days old with an interval of at least 1h between test sessions each day. We analyzed sessions in which one experimenter presented to infants a lipsmacking gesture. In each test session, one experimenter held the infant swaddled in pieces of fleece fabric. A second experimenter served as the source of the stimuli, and a third experimenter videotaped the test session (30 frames per second, using a Sony Digital Video camcorder ZR600) and ensured correct timing of the different phases of the trial. At the beginning of a trial, a 40 sec baseline was conducted, in which the demonstrator displayed a passive/neutral facial expression. The demonstrator then displayed the lipsmack gesture for 20 seconds, followed by a still face period for 20 sec. This stimulus-still face sequence was repeated three times.

Chewing behavior was not recorded from neonatal monkeys because they do not eat solid food in the early weeks of postnatal life.

Juvenile and adult lipsmack data collection

Juvenile and adult lipsmacks were recorded from subjects on the island of Cayo Santiago, Puerto Rico. Cayo Santiago is home to a semi-free ranging colony of approximately 1100 rhesus monkeys, funded and operated jointly by the National Institutes of Health and the University of Puerto Rico. Monkeys are provisioned daily with monkey biscuits. Subjects recorded were both male and female monkeys, classified as either juveniles (6 mos. to ~3.5 yrs, n=16; Figure 1B) or adults (older than 3.5 yrs, n=22; Figure 1C). In the adult age class, lipsmacks were recorded primarily from females due to the affiliative nature of the display and the aggressive tendencies of adult males in the field (Maestripieri & Wallen 1997). Lipsmacks from the juvenile age class were recorded from both males and females. Observed behaviors in Cayo Santiago included lipsmacks and chewing. All video data were collected with a Canon Vixia HF100 camcorder at 30 frames per second. Lipsmacks were elicited through interactions with human experimenters opportunistically engaging in eye contact with individuals on the island.

Chewing behavior in both adults (n=10) and juveniles (n=10) was recorded *ad libitum* during ingestion of provisioned monkey biscuits. Only chewing of these biscuits was recorded to avoid potential influence of factors such as food type or hardness on rate of mastication. Chewing bouts of both males and females in each age class were included. Video recording of chewing bouts was performed as described above for lipsmacks.

Extracting temporal dynamics

Lipsmack and chewing video clips were analyzed using MATLAB (Mathworks, Natick, MA) for vertical mouth displacement as a function of time (Figure 2). All videos were 30 frames per second. The Nyquist frequency, above which temporal dynamics cannot be reliably extracted, is thus 15 Hz ($f_{Nyquist} = \frac{1}{2}$ v, where v = sampling rate). Mouth

displacement was measured frame-by-frame. For lipsmacks and chewing, 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 lipsmacks, the top and bottom lips do not part, so inter-lip distance does not provide a reliable indication of jaw displacement. For these lipsmacks, 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.

Spectral analysis of mouth oscillatory rhythms

To quantify the rhythms of each mouth displacement time series, spectral analysis was performed with a multi-taper Fourier transform (Chronux Toolbox, www.chronux.org). Because of the Nyquist limit frequency, the band pass was set as $0 \leq f_{pass} \leq 15$ Hz. A power spectrum was generated for each bout and the peak was measured in MATLAB. This peak reflects the periodicity which is most representative of a given time series, and was considered to be the approximate rate of mouth oscillation in each bout. The mouth opening time series were amplitude normalized so that maximum mouth opening in every time series was uniform, preventing variability in power solely as an artifact of mouth opening distance. Amplitude normalization of time series does not affect the location of peaks in the frequency domain, but does allow for comparison of mouth oscillation spectra across multiple bouts without biasing for differences in mouth opening size.

Statistical analysis

Independent *t*-tests were performed. When necessary, we accounted for unequal variances and Bonferroni-corrected p-values. All statistical analyses were performed with the MATLAB Statistics Toolbox or SPSS Statistics 17.0. We also calculated the coefficient of variation (CV) for lipsmacks and chewing behavior. The CV is a dimensionless ratio of standard deviation (SD) over the mean, thereby describing the variance in the context of a mean value. For example, a constant SD results in a larger CV as mean decreases. This measure is thus appropriate when comparing distributions in which means vary significantly.

Results

If visuofacial gestures such as monkey lipsmacks represent an evolutionary precursor to the rhythmic structure of speech, then we hypothesized that lipsmacks and speech would exhibit parallels in temporal dynamics and developmental trajectories. We video-recorded the lipsmacking of rhesus monkeys (*Macaca mulatta*) from three different age groups: neonates (3–8 days old), juveniles (6 months to 3.5 years old) and adults (greater than 3.5 years of age)(Figure 1A–C). We then extracted the temporal dynamics of this rhythmic mouth movement and performed spectral analyses on them (Figure 2). We first measured the lipsmack frequency in adult monkeys for comparison with that of other age classes (n=16; Figure 3A,B). Adult lipsmack rhythm was 5.14±0.82 Hz (mean ± SD)(Figure 3C), similar to what we reported previously (Ghazanfar et al. 2010). This resembles the ~5Hz rhythm of syllable production in human speech (Chandrasekaran et al. 2009; Crystal & House 1982; Dolata et al. 2008; Greenberg et al. 2003; Malecot et al. 1972). The coefficient of variation (CV) for adult lipsmacks was 0.16 (Figure 3D).

We next examined the rhythmic frequency of lipsmacks produced by neonatal monkeys between 3 and 8 days of age (n=15; Figure 4A,B). Neonatal lipsmacks showed both relatively lower frequencies and greater variability when compared to adults. The mean neonatal lipsmack rhythm was 1.52±0.94 Hz (Figure 4C), which is significantly different from mean adult rhythm (two-sample t-test (t(29) = 12.43, corrected p < 0.0001). The CV

for neonatal lipsmacks was 0.62 (Figure 4D), almost four times greater than the adult variability. This pattern is similar to human neonatal babbling which is characterized by a lower mean frequency (Dolata et al. 2008; Levitt & Wang 1991; Lynch et al. 1995; Nathani et al. 2003) and high variability (Dolata et al. 2008) relative to mature speech.

We then measured the rhythmic frequency of juvenile monkey lipsmacks; these were produced by monkeys between the ages of 6 months and 3.5 years (n = 22) (Figure 5A,B). Our data show that the juvenile lipsmack has an adult-like mean frequency, but with greater variability. Juvenile lipsmack frequencies showed a mean of 5.33±1.69 Hz (Figure 5C), which is not significantly different from the mean of adult frequencies (t(36) = 0.50, n.s.). It is, however, significantly different from the neonatal lipsmack rhythm (t(35) = 8.75, corrected p < 0.0001). Comparing the variability of the juvenile lipsmacks (CV=0.32) to adult (CV=0.16) and neonatal (CV=0.62) lipsmacks shows that rhythmic variability decreases by almost half at each stage of development. This differential variability is apparent when the mean power spectra of lipsmack oscillations for juveniles and adults are compared (Figures 3B and 5B). While the mean adult spectrum shows a clear peak, the juvenile spectrum plateaus in the approximate range of 3–8 Hz, which extends both above and below the peak adult range.

Overall, a comparison of lipsmacks across the three age groups suggests lipsmacks go through multiple stages of development. There is an increase in rhythmic frequency up to ~5 Hz from neonatal to juvenile stages. From the juvenile stage onward, the maturation required to reach adult-like lipsmacks appears to be a refinement of control (decrease in variability) rather than a directional change in mean frequency.

One possibility is that the changes in lipsmack temporal dynamics over the course of development are the outcome of a more general maturation of orofacial motor circuits. Indeed, it would not be unreasonable to expect that any complex movement results from changes that go from slow and variable to faster and less variable. To test for this, we analyzed temporal dynamics of another rhythmic mouth movement, that of chewing, which uses the same facial anatomical structures as lipsmacking. We did this for both juvenile (n=10; Figure 6A,B) and the adult monkeys (n=10; Figure 6E,F). [We did not measure neonatal chewing movements, as they do not exist; rhesus monkeys do not eat solid food in the first weeks of life]. Juvenile mean frequency of chewing was 2.49±0.34 Hz (Figure 6C), while adult mean frequency was 2.60 Hz±0.34 Hz (Figure 6G). These means are not significantly different from each other (t(18) = 0.68, n.s.). This rhythmic frequency for chewing is similar to that of humans. When compared to lipsmacks, the rhythmic frequency of chewing was significantly slower in both age groups (juvenile: t(30) = 7.49, p < 0.0001; adult: t(24) = 13.15 p < 0.0001). Also in contrast to lipsmack development, the variability of chewing movements did not change much between juveniles and adults (juvenile CV = 0.14; adult CV = 0.13)(Figure 6D,H). Overall, these data show that the development dynamics of lipsmacking do not represent a general change in the degree of orofacial motor control, but a refinement of a distinct circuit.

Another possibility is that the rhythmic frequency of lipsmacking could be different in captivity (where our neonatal lipsmacks were collected) versus semi-free-ranging conditions. Though we couldn't examine this issue for all age groups, we compared the lipsmacking by adult macaques on Cayo Santiago with adult macaques in captivity while they were seated in monkey chairs. Lip-smacking is at 5.14±0.82 Hz for wild monkeys (as reported above) and 5.82±0.90 Hz for captive monkeys. There were no meaningful differences.

Discussion

Our hypothesis from the outset was that, if the rhythmic nature of audiovisual speech evolved from the rhythmic facial expressions of ancestral primates (MacNeilage 1998; MacNeilage 2008), then macaque monkey lipsmacks should not only have the same rhythmic frequency (Ghazanfar et al. 2010), but should also develop with the same trajectory as human speech. We measured the rhythmic frequency and variability of lipsmacks across individuals in three different age groups: neonates, juveniles and adults. There were many possible outcomes. First, given the differences in the size of the facial structures between macaques and humans, there was a distinct possibility that lipsmacks and speech rhythms need not converge on the same ~5 Hz rhythm. Second, because of the precocial neocortical development of macaque monkeys relative to humans (Gibson 1991; Malkova et al. 2006), the lipsmack rhythm could remain stable from birth onwards, showing no changes in frequency and/or no changes in variability (Thelen 1981). Finally, whatever changes in lipsmack structure may occur, they do not have to diverge from that of chewing, another rhythmic mouth movement. That is, it may be that all complex motor acts undergo the same changes in frequency and variablity (for example, rhythmic limb movements in humans infants (Thelen 1979)).

In light of these alternative possibilites, it is striking that our data show that lipsmacking develops like the human speech rhythm: young individuals produce slower, more variable mouth movements and as they get older, these movements become faster and less variable. Furthermore, as in human speech development (Smith & Zelaznik 2004), the variability and frequency changes in lipsmacking are independent in that juveniles have the same rhythmic lipsmack frequency as adult monkeys, but the lipsmacks are much more variable. In juvenile and adult monkeys, the mean rhythmic frequency is ~5 Hz frequency, the same frequency that characterizes the speech rhythm in human adults (Chandrasekaran et al. 2009; Crystal & House 1982; Dolata et al. 2008; Greenberg et al. 2003; Malecot et al. 1972). Importantly, the developmental trajectory for lipsmacking was different from that of chewing. Chewing had the same slow frequency (~2.5 Hz) as in humans and consistent low variability across the juvenile and adult age groups. These differences in developmental trajectories between lipsmacking and chewing are identical to those reported in humans for speech and chewing (Moore & Ruark 1996; Steeve 2010; Steeve et al. 2008). Naturally, as Old World monkeys develop precocially relative to humans (Gibson 1991; Malkova et al. 2006), these parallel trajectories occur on different timescales (i.e, faster in macaques) (Clancy et al. 2000; Kingsbury & Finlay 2001). Overall, our data suggest that monkey lipsmacking and human speech share a homologous developmental mechanism and lend strong empirical support for the idea that human speech evolved from the rhythmic facial expressions of our primate ancestors (MacNeilage 1998; MacNeilage 2008).

What about the lack of vocal output in lipsmacking?

Lipsmacks may be the most ubiquitous of all the primate facial expressions, occurring in a wide range of social contexts and observed in many genera of Old World primates (Redican 1975; Van Hooff 1962; Goodall 1968; Parr et al. 2005). In the present study, we've shown that not only is the lipsmack rhythmic frequency the same as adult audiovisual speech (Chandrasekaran et al. 2009), but its developmental trajectory is also similar to the one that takes humans from babbling in infancy to adult speech. Yet, a fundamental difference between lipsmacking and babbling/speech is that the former lacks a vocal (acoustic) component. Thus, the capacity to coordinate vocalization during cyclical mouth movements seen in babbling and speech seems to be a human adaptation. How can one reconcile this difference? That is, how can lipsmacks be related to speech if there is no vocal component?

In human and nonhuman primates, the basic mechanics of voice production are broadly similar and consists of two distinct components: the source and the filter (Fant 1970; Fitch & Hauser 1995; Ghazanfar & Rendall 2008). Voice production involves 1) a sound generated by air pushed by the lungs through the *larynx* (the source) and 2) the modification through resonance of this sound by the *vocal tract* airways above the larynx (the filter: the nasal and oral cavities whose shapes can be changed by movements of the jaw, tongue and lips). These two basic components of the vocal apparatus behave and interact in complex ways to generate a wide range of sounds; the movements of the mouth are essential and intimately linked to the acoustics generated by the laryngeal source, including its rhythmic structure (Chandrasekaran et al. 2009). Thus, MacNeilage's theory (MacNeilage 1998; MacNeilage 2008)--purporting that the origins of rhythmic speech occurred via modifications of rhythmic facial gestures--addresses only the vocal tract filter component of babbling/speech production. In support of MacNeilage's idea, our developmental data show that the rhythmic facial movements required by human speech likely have their origin in the lipsmacking gesture of nonhuman primates. The separate origin of laryngeal control (that is, how to link the voice with the rhythmic facial gestures) remains a mystery. The most plausible scenario is that the cortical control of the brainstem's *nucleus ambiguus*, which innervates the laryngeal muscles, is absent in all primates save for humans (Deacon 1997).

Different orofacial behaviors have different developmental trajectories

We found that juvenile and adult lipsmack frequencies were faster than chewing movements. Furthermore, from the juvenile stage to adults, the developmental changes were different for lipsmacks versus chewing. At the juvenile stage, lipsmacks were highly variable relative to adult lipsmacks, while chewing movements had consistent low variability between these two age groups. These differences in the pattern of development for lipsmacks versus chewing are similar to those seen in humans for chewing and speech. Like macaque monkey chewing, the frequency of chewing movements in human infants is highly stereotyped, remaining virtually unchanged from early infancy into adulthood (Green et al. 1997; Kiliaridis et al. 1991). In contrast, for both monkey lipsmacking and human speech, the pattern of change is very different. As we show for lipsmacking, the human infant babbling rhythm increases in frequency and decreases in variability as it transitions to adult speech (Green et al. 2002; Smith & Goffman 1998; Tingley & Allen 1975). These differences in the developmental trajectories for chewing versus lipsmack and speech suggest that different underlying neurophysiological mechanisms are at play.

The mandibular movements shared by chewing, lipsmacking 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). While chewing movements may be largely independent of cortical control (Lund & Kolta 2006), lipsmacking and speech production are both modulated by the neocortex, in accord with social context and communication goals (Bohland & Guenther 2006; Caruana et al. 2011). Thus, one hypothesis is that the developmental changes in the frequency and variability of lipsmacks and speech are a reflection of the maturation of neocortical circuits influencing brainstem central pattern generators (as suggested by Thelen (Thelen 1981) in the context of rhythmic movements more generally).

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 uncertain social environments (Preuschoff et al. 2008). Importantly, the human insula is also involved in speech production (Ackermann & Riecker 2004; Catrin Blank et al. 2002; Dronkers 1996); (Bohland & Guenther 2006). Consistent with a evolutionary link between lipsmacks and speech, the insula also plays a role in generating monkey lipsmacks (Caruana et al. 2011). Electrical stimulation of the insula elicits lipsmacking 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 lipsmack production. Thus, it is conceivable that for the both monkey lipsmacking and human speech, the increase in rhythmic frequency and decrease in variability are, in part at least, due to 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 lipsmacks (Ferrari et al. 2003).

Does social context influence the development of both lipsmacking and babbling?

Monkey lipsmacking and human infant babbling show a high degree of variability in their rhythmic frequency that gets reduced over time. Typically, high variability would be associated with immature motor neuronal function. Indeed, Thelen (1981) has suggested that high motor variability early in life is associated with the slow myelination process in cortical circuits and that this may be adaptive: the high variability is useful for providing a flexible substrate that can allow social input to guide further development. For example, children need flexibly organized motor control systems so that they can acquire new patterns of speech. The formation of those new patterns is often guided by feedback and instruction from caregivers. Human infants given contingent feedback from their mothers rapidly re-structure their babbling output and do so in accordance to the caregiver's generic, attentive responses as well as towards the caregivers' specific phonological output (Goldstein et al. 2003; Goldstein & Schwade 2008).

This type of social feedback is not limited to the humans. Mother-infant pairs of macaque monkeys also communicate inter-subjectively (Ferrari et al. 2009). They exchange complex forms of communication that include mutual gaze, mouth-to-mouth contacts, imitation and, most important for the present discussion, lipsmacks. Lipsmacking between mothers and neonatal infant monkeys always require mutual gaze. Although unrelated adult monkeys exchange lipsmacks, some patterns of lipsmacks between mothers and neonates are unique (Ferrari et al. 2009). For example, sometimes a mother holds the infant's head, pulling its face towards her own, before producing lipsmacks. Other times, a mother moves her head up and down, close to the infant, in order to capture its attention before producing lipsmacks. Could these maternal lipsmacks guide infant monkeys to produce lipsmacks with a consistent species-typical rhythmic frequency?

The evolution of speech

MacNeilage (MacNeilage 1998; MacNeilage 2008) suggested that during the course of human speech evolution, rhythmic facial expressions were coupled to vocalizations. We tested one aspect this prediction by investigating the structure and development of macaque monkey lipsmacks and found that their developmental trajectory is strikingly similar to the one that leads from human infant babbling to adult speech: younger monkeys produce slower, more variable mouth movements and as they get older, these movements become faster and less variable. This developmental pattern does *not* occur for another cyclical mouth movement—chewing. Ultimately, both lipsmacking and speech converge on a stable ~5 Hz rhythm that represents the frequency that characterizes the average syllable production rate in human adults (Chandrasekaran et al. 2009; Crystal & House 1982; Dolata et al. 2008; Greenberg et al. 2003; Malecot et al. 1972). We suggest that monkey lipsmacking and human babbling share a homologous developmental mechanism, lending

strong empirical support for the idea that human speech evolved from the rhythmic facial expressions of our primate ancestors (MacNeilage 1998; MacNeilage 2008).

Acknowledgments

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]

Bell, A.; Hooper, JB. Syllables and segments. Amsterdam: North-Holland; 1978.

Bohland JW, Guenther FH. An fMRI investigation of syllable sequence production. Neuroimage. 2006; 2:821–841. [PubMed: 16730195]

Buzsaki G, Draguhn A. Neuronal oscillations in cortical networks. Science. 2004; 304:1926–1929. [PubMed: 15218136]

Campbell R. The processing of audio-visual speech: empirical and neural bases. Philosophical Transactions of the Royal Society B: Biological Sciences. 2008; 363:1001–1010.

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, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA. The natural statistics of audiovisual speech. PLoS Computational Biology. 2009; 5:e1000436. [PubMed: 19609344]

Clancy B, Darlington RB, Finlay BL. The course of human events: predicting the timing of primate neural development. Developmental Science. 2000; 3:57–66.

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]

Deacon TW. Rethinking mammalian brain evolution. American Zoologist. 1990; 30:629–705.

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. J Acoust Soc Am. 1994; 95:2670–80. [PubMed: 8207140]

Fant, G. Acoustic theory of speech production. Paris: Mouton; 1970.

Ferrari P, Visalberghi E, Paukner A, Fogassi L, Ruggiero A, Suomi S. Neonatal imitation in rhesus macaques. PLoS Biology. 2006; 4:1501.

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 S. Reciprical face-to-face communication between rhesus macaque mothers and their newborn infants. Current Biology. 2009; 19:1768–1772. [PubMed: 19818617]

Finlay BL, Darlington RB, Nicastro N. Developmental structure of brain evolution. Behavorial and Brain Sciences. 2001; 24:263–308.

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, 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, Rendall D. Evolution of human vocal production. Curr Biol. 2008; 18:R457–R460. [PubMed: 18522811]

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, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron. 2007; 56:1127–1134. [PubMed: 18093532]

Goldstein MH, King AP, West MJ. Social interaction shapes babbling: Testing parallels between birdsong and speech. Proc Natl Acad Sci, USA. 2003; 100:8030–8035. [PubMed: 12808137]

Goldstein MH, Schwade JA. Social feedback to infants' babbling facilitates rapid phonological learning. Psychological Science. 2008; 19

Goodall, J. A preliminary report on expressive movements and communication in the Gombe Stream chimpanzees. In: Jay, PC., editor. Primates: Studies in Adaptation and Variability. New York: Holt, Rinehart and Winston; 1968. p. 313-519.

Gottlieb, G. Individual development & evolution: the genesis of novel behavior. New York: Oxford University Press; 1992.

Green JR, Moore CA, Reilly KJ. The sequential development of jaw and lip control for speech. Journal of Speech, Language and Hearing Research. 2002; 45:66–79.

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.

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.

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]

Kim J, Davis C. Investigating the audio-visual speech detection advantage. Speech Communication. 2004; 44:19–30.

Kingsbury MA, Finlay BL. The cortex in multidimensional space: where do cortical areas come from? Developmental Science. 2001; 4:125–157.

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]

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: Kluwer Academic Publishers; 1996.

Locke, JL. The child's path to spoken language. Cambridge, MA: Harvard University Press; 1993.

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]

Luo H, Liu Z, Poeppel D. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biology. 2010; 8:e1000445. [PubMed: 20711473]

Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 2007; 54:1001–1010. [PubMed: 17582338]

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. [PubMed: 10097020]

MacNeilage, PF. The origin of speech. Oxford, UK: Oxford University Press; 2008.

Maestripieri D, Wallen K. Affiliative and submissive communication in rhesus macaques. Primates. 1997; 38:127–138.

Malecot A, Johonson R, Kizziar PA. 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]

Moore CA, Ruark JL. Does speech emerge from earlier appearing motor behaviors? Journal of Speech and Hearing Research. 1996; 39:1034–1047. [PubMed: 8898256]

Nathani S, Oller DK, Cobo-Lewis A. Final syllable lengthening (FSL) in infant vocalizations. Journal of Child Language. 2003; 30:3–25. [PubMed: 12718291]

Oller, DK. The emergence of the speech capacity. Mahwah, NJ: Lawrence Erlbaum; 2000.

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.

Pinker S, Bloom P. Natural language and natural selection. Behavioral and Brain Sciences. 1990; 13:707–784.

Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication. 2003; 41:245–255.

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.

Ruppenthal GC, Arling GL, Harlow HF, Sackett GP, Suomi SJ. A 10-year perspective of motherless-mother monkey behavior. Journal of Abnormal Psychology. 1976; 85:341–349. [PubMed: 821983]

Saberi K, Perrott DR. Cognitive restoration of reversed speech. Nature. 1999; 398:760–760. [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.

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. Neuronal oscillations and visual amplification of speech. Trends Cogn Sci. 2008; 12:106–113. [PubMed: 18280772]

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech Recognition with Primarily Temporal Cues. Science. 1995; 270:303–304. [PubMed: 7569981]

Smith A, Goffman L. Stability and patterning of speech movement sequences in children and adults. Journal of Speech, Language and Hearing Research. 1998; 41:18–30.

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]

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.

Thelen E. Rhythmical Stereotypies in Normal Human Infants. Animal Behaviour. 1979; 27:699–715. [PubMed: 556122]

Thelen E. Rhythmical Behavior in Infancy - an Ethological Perspective. Developmental Psychology. 1981; 17:237–257.

Tingley BM, Allen GD. Development of speech timing control in children. Child Development. 1975; 46:186–194.

Van Hooff JARAM. Facial expressions of higher primates. Symposium of the Zoological Society, London. 1962; 8:97–125.

Vitkovitch M, Barber P. Effect of Video Frame Rate on Subjects' Ability to Shadow One of Two Competing Verbal Passages. J Speech Hear Res. 1994; 37:1204–1210. [PubMed: 7823562]

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.

Figure 1. Lipsmack frame-by-frame exemplars from neonatal, juvenile and adult monkeys
A, Neonatal lipsmack exemplar, consisting of one open-close alternation. The lips begin closed and slightly puckered, separate and close again to a pucker. This constitutes one cycle. Tongue protrusion, seen clearly in the middle frame, is a common feature of the lipsmack. **B,** Juvenile lipsmack, one cycle. **C,** Adult lipsmack, one cycle.
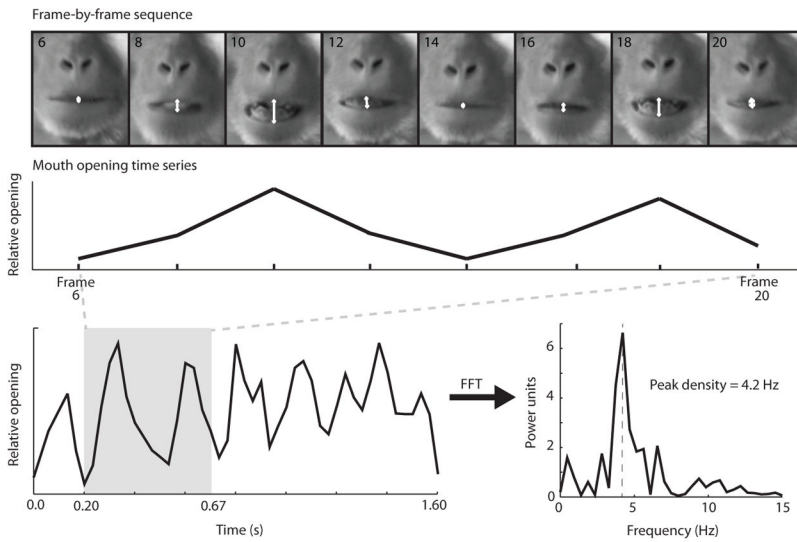
Figure 2. Manual coding of inter-lip distance and time series analysis
Methods used to extract temporal dynamics from video sequences of the target orofacial gesture. The frame-by-frame sequence shows ~2 open-close alternations during a lipsmack performed by an adult monkey. Manual coding was performed using a custom computer program by indicating a point in the middle of the upper lip and a point in the middle of the lower lip for each frame. White arrows show approximate points of mouse click for this sequence, and the white line between them shows inter-lip distance. This inter-lip distance is then converted into a time series of mouth opening through time, which is calculated from the frame rate of each video clip (30 fps). The longer time series on the bottom right represents the complete lipsmack bout of which this 2-cycle sequence is part. To determine the most representative frequency of the time series, a fast Fourier transform (FFT) is performed as described in the methods and a power spectrum is generated. The frequency at which peak spectral density occurs is considered the most representative frequency of the gestural bout.
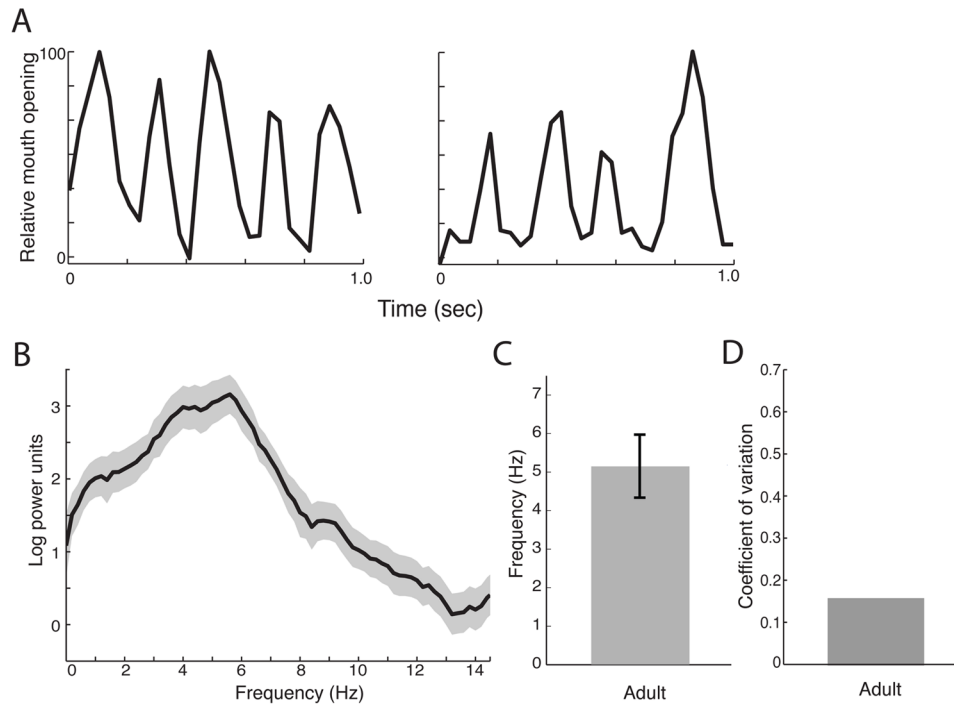
Figure 3. Rhythmic dynamics of adult monkey lipsmacks
A, Exemplar time series, showing 1 s samples of adult lipsmack mouth oscillation at frequencies representative of data set. Vertical axis represents normalized units for mouth opening, with peak inter-lip distance during the sequence set at 100. **B,** Mean power spectrum for all adult lipsmack bouts consisting of power spectra from all adult lipsmack data averaged. Shaded area above and below spectrum line indicates ± 1 SEM. **C,** Mean of peak spectral densities from each adult lipsmack bout (mean±SD). Error bars indicate ± 1 SD **D**, Coefficient of variation for adult lipsmacks.
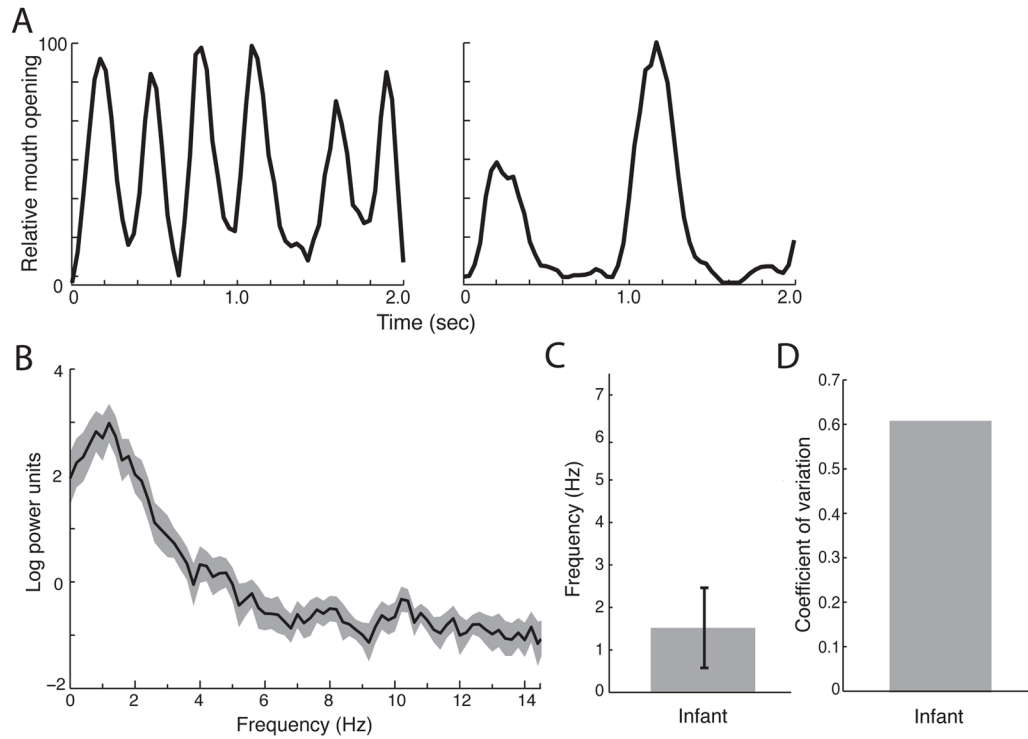
Figure 4. Rhythmic dynamics of neonatal monkey lipsmacks
A, Exemplar time series, showing 1 s samples of mouth oscillation at frequencies representative of data set. Axes of all plots as in Figure 3. **B,** Mean power spectrum for all neonatal lipsmacks bouts. **C,** Mean of peak spectral densities from each neonatal lipsmack bout (mean±SD). Error bars indicate ± 1 SD. **D,** Coefficient of variation for neonatal lipsmacks.

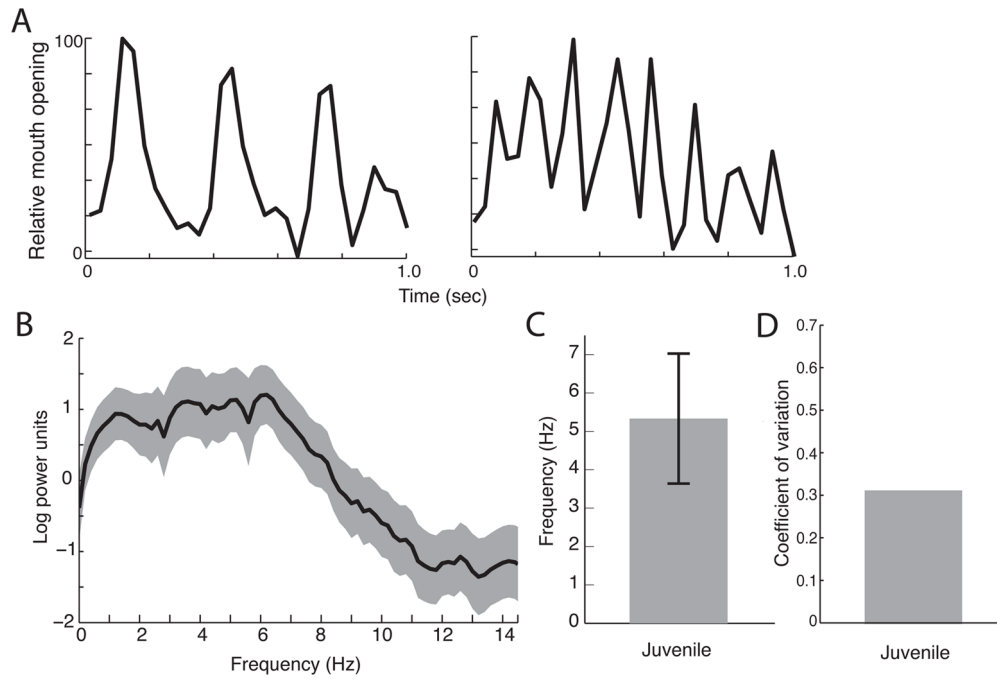**Figure 5. Rhythmic dynamics of juvenile monkey lipsmacks**
A, Exemplar time series, showing 1 s samples of mouth oscillation at frequencies representative of data set. Axes of all plots as in Figure 3. **B,** Mean power spectrum for all adult lipsmack bouts. **C,** Mean of peak spectral densities from each juvenile lipsmack bout (mean±SD). Error bars indicate ± 1 SD. **D**, Coefficient of variation for juvenile lipsmacks.
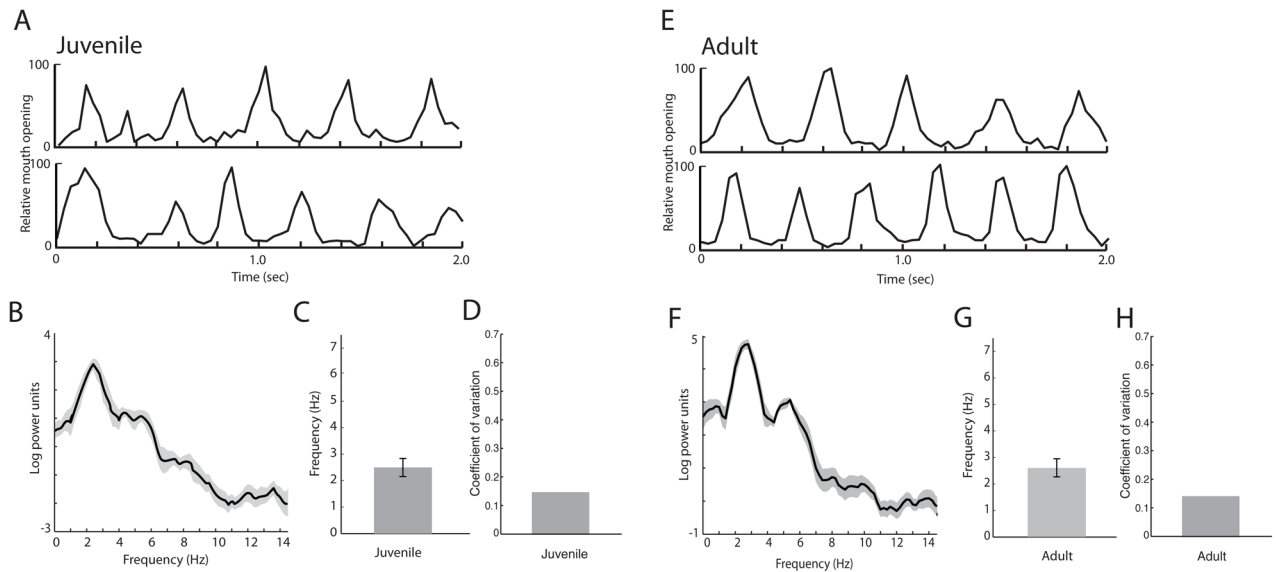
Figure 6. Chewing rhythmic dynamics between juvenile and adult monkeys
A–D, Temporal dynamics of juvenile monkey chewing. **A,** Two exemplar juvenile chewing time series of 2 s length. Vertical axis represents normalized units for mouth opening, with peak inter-lip distance during the sequence set at 100. **B,** Juvenile chewing mean power spectrum, consisting of power spectra of all juvenile chewing bouts averaged. **C,** Mean of peak spectral densities from each juvenile lipsmack bout (mean±SD) Error bars indicate ± 1 SD. **D**, Coefficient of variation for juvenile chewing movements. **E–H,** Temporal dynamics of adult chewing. **E,** Two 2 s exemplars of adult chewing time series. **F,** Adult chewing mean power spectrum. **G,** Mean of peak spectral densities from each adult chewing bout (mean±SD). **H**, Coefficient of variation for adult chewing movements.