Imitation of Non-Speech Oral Gestures by 8-Month-Old Infants

Heidi Diepstra  
Department of Speech and Language Pathology, University of Toronto, Canada

Sandra E Trehub  
Department of Psychology, University of Toronto Mississauga, Canada

Alice Eriks-Brophy  
Department of Speech and Language Pathology, University of Toronto, Canada

Pascal HHM van Lieshout  
Department of Speech and Language Pathology, University of Toronto, Canada

Abstract
This study investigates the oral gestures of 8-month-old infants in response to audiovisual presentation of lip and tongue smacks. Infants exhibited more lip gestures than tongue gestures following adult lip smacks and more tongue gestures than lip gestures following adult tongue smacks. The findings, which are consistent with predictions from Articulatory Phonology, imply that 8-month-old infants are capable of producing goal-directed oral gestures by matching the articulatory organ of an adult model.

Keywords
Infants, babbling, imitation, oral gestures, phonology

Introduction
Babbling continues to intrigue scholars because it provides a glimpse into the most primitive units of speech. The production of speech syllables and canonical babbling (e.g., “bababa”) begins at approximately 6–8 months of age (Oller, 1980; Stark, 1980). Infants succeed in producing some of the speech sounds in their environment, but the units of information that guide them toward this perceptual-motor achievement remain unknown. Classic as well as some more contemporary formalist linguistic theories of the acquisition of speech (e.g., Boersma & Levelt, 2003; Chomsky &...
Halle, 1968; Jakobson & Halle, 1956) describe development of the speech sound system (phonology) based on language rules and markedness constraints, with more attention to the development of language competence than to performance. In this view, a priori knowledge of language form is assumed, and phonological development is considered the unfolding of a set of innate rules. In other words, formal phonological approaches are relatively unconcerned with the origin of generative rules or constraints, either in ontogeny or phylogeny. In addition, they pay little or no attention to the motor aspects of speech development.

Contrary to formalist approaches, functionalist or emergence theories of phonological development reject the notion of pre-existing language knowledge that guides a child to speech development (e.g., Davis & Bedore, 2013; Goldstein & Fowler, 2003; Kent, 1984; Lindblom, 2000; Thelen, 1991). Instead, speech is considered to emerge from an interaction of the infant’s perceptual, cognitive and motor systems, which develops in the context of the social and physical environment. From this perspective, phonological development is mainly considered in terms of perceptual-motor constraints. Recent advancements in computational modeling techniques contributed to the popularity of two emergence approaches: The Frame-then-Content theory (FC; MacNeilage & Davis, 2000) and Articulatory Phonology (AP; Browman & Goldstein, 1989, 1992). Both theories consider speech as motor action in time, with phonological structure arising from biomechanical characteristics and constraints of the vocal tract rather than from an innate language module (Chomsky, 1986). They differ, however, in assumptions about the control structures underlying the emergence of syllables in babbling and first words (Giulivi, Whalen, Goldstein, Nam, & Levitt, 2011; MacNeilage & Davis, 2011; Whalen, Giulivi, Goldstein, Nam, & Levitt, 2011). A better understanding of the mechanisms underlying oral motor control in babbling is necessary to advance emergence approaches to speech acquisition.

In this study, we use an imitation task with two target behaviors to investigate contrasting predictions from AP and FC regarding articulator control in babbling. According to AP, the most primitive units of phonology are constrictions of vocal tract organs (i.e., lips, tongue tip, tongue body, tongue dorsum, velum and glottis), also called gestures (Browman & Goldstein, 1989; Goldstein, Byrd, & Saltzman, 2006). The actions of gestural constriction are defined as abstract tasks that are specified by dynamic parameters (stiffness, damping, mass) controlling constriction location and degree. In other words, gestures are goal-directed actions implemented by sets of articulators, or synergies. (Saltzman & Munhall, 1989). For example, the lip closure at the start of the word “pan” involves actions of the lips and jaw. Specific phase relationships define the coordination or relative timing of gestures into larger structures or utterances (e.g., syllables, words). This intergestural timing involves a planning process that takes lexical specifications of pre-determined phase values as input for a context-specific constellation of gestural actions (Saltzman & Byrd, 2000).

Gestures must be discrete and combinable to function as basic phonological units (Goldstein, 2003). One source of discreteness is the independent constriction of different articulatory organs in the vocal tract creating between-organ contrasts. For example, onsets of the words “tan” and “pan” contrast in the use of a tongue tip or lip gesture. Same-organ gestures may also differ in their parameter settings for constriction location and degree, providing a second source of discreteness. The utterances “thick”, “sick” and “tick” vary in the location or degree of constriction for the same organ, namely the tongue tip. Such within-organ differentiation of gestures is assumed to emerge through progressive attunement of production to sounds in the environment, as suggested by several modeling studies (Browman & Goldstein, 2000; De Boer, 2000; Goldstein, 2003).

Infants are born with discrete vocal tract organs and the development of within-organ phonological contrasts requires interaction with language users. Therefore, AP considers between-organ contrasts to be the most primitive phonological contrasts (Studdert-Kennedy, 2000). Studies of facial mimicry in infancy provide support for this hypothesis. For example, when an adult model
produces mouth opening or tongue protrusion, newborn infants respond with the correct organ, whether lips or tongue (Meltzoff & Moore, 1997). The organ hypothesis in AP implies that infants have access to distinct articulatory synergies with some control of tongue and lips independent from the jaw, enabling the infant to perform goal-directed actions of constriction. Thus, even though infants lack adult-like control of individual consonant and vowel gestures, opening and closing movements of the jaw may involve additional oscillation of other organs; for example, lips for a consonant constriction or tongue body for a vowel gesture (Nam, Goldstein, Giulivi, Levitt, & Whalen, 2013). Infants are presumed capable of coordinating tongue and lips along with the jaw because of well-established feeding behaviors that require some degree of independent control of lips and tongue (Nam et al., 2013). From this perspective, one would predict infant responding with the correct articulatory organs when presented with speech or non-speech oral gestures such as lip smacks and tongue smacks in audiovisual and, because most of the vocal tract organs are not visible during speech (e.g., velum, glottis), in auditory-only contexts (Goldstein, 2003; Studdert-Kennedy & Goldstein, 2003). Infants are presumed to do so because they use the same unit for speech perception and production, which is the vocal tract action or gesture (Goldstein & Fowler, 2003). To date, however, there has been no empirical evidence of imitation of such oral gestures by infants in audiovisual or auditory-only contexts.

Like AP, the FC considers babbling as an oscillatory action, originating in the cyclic mandibular actions related to ingestive behaviors such as chewing (Davis & MacNeilage, 1995; MacNeilage, 1998). Oral constrictions in the closing phase of the jaw oscillation and subsequent jaw aperture accompanied by phonation are perceived as consonants and vowels, respectively. In FC, the mandibular oscillatory pattern of movement provides a frame for syllables in babbling (Davis & MacNeilage, 1995; MacNeilage, 1998). Certain consonant-vowel (CV) combinations occur at greater than chance levels in babbling and first words (Davis & MacNeilage, 1995; Davis, MacNeilage, & Matyear, 2002). Infants prefer labial-central, coronal-front and dorsal-back CV combinations (MacNeilage & Davis, 2000), which have been noted in different languages (Giulivi et al., 2011), in dictionary data (MacNeilage, Davis, Kinney, & Matyear, 2000) and, to some extent, in adult corpora (Whalen et al., 2012). According to FC, random positioning and inertia of the tongue during jaw oscillation provide a biomechanical account of these CV biases. Fundamental to FC theory is infants’ lack of independent control of their articulators until they have sufficient motor control to differentiate segmental content within the syllable-frame dominance (MacNeilage, 1998). FC theory was the only biomechanical account of CV jaw-tongue combination preferences until recent questions arose about its plausibility (MacNeilage & Davis, 2011; Nam et al., 2013). The jaw-only control hypothesis in FC implies that increased articulatory control diminishes CV biases, but no such changes were observed in a longitudinal study of 6-month to 8-month-old babblers (Giulivi et al., 2011). Because adults have independent control over tongue and lips, CV biases of the kind observed in infants should disappear in their lexicons, but dictionary counts indicate the persistence of CV preferences (Giulivi et al., 2011). Finally, syllables with the predicted CV bias in babbling account for about half of infants’ babbled utterances (Giulivi et al., 2011). If babble was dominated by jaw-only action, the ratio of observed-to-expected CV combinations would be much larger for the preferred combinations than for the non-preferred combinations (Giulivi et al., 2011).

AP offers an alternative biomechanical account of the CV bias in babbling and adult language (Goldstein et al., 2006). In addition to jaw control, AP proposes that infants have some control of vocal tract constriction by tongue, lips and velum. Moreover, planning oscillators for different gestures can be coupled in-phase or anti-phase, as in any other oscillating systems (Haken, Kelso, & Bunz, 1985). For CV syllables in particular, consonant and vowel gestures are coupled in a stable in-phase mode (Goldstein et al., 2006), which means that both gestures are triggered simultaneously within the biomechanical constraints of the vocal tract. Consonant and vowel gestures that
are inherently compatible are either mechanically independent (lips and tongue as in /ba/) or constricted at compatible locations (coronal with front part of tongue, and dorsal with back part of tongue). According to AP, then, certain vowel gestures synchronize more easily than others with consonant gestures. For /di/ the tongue body takes an advanced position for both consonant and vowel, while for /du/ tongue body actions in opposite directions are required. This gestural account explains CV co-occurrence preferences in both babbling and adult speech (Goldstein et al., 2006).

In summary, AP and FC perspectives concur on the role of biomechanical factors in the emergence of babbling but disagree on the role of underlying control structures. AP argues that because of movement synergy, tongue body constrictions that coordinate in a naturally stable way with consonant constrictions will emerge spontaneously. By contrast, FC posits that CV biases are due to jaw-only control. Computational simulations reveal better agreement with patterns of CV preferences in infant babbling based on AP than on FC (Nam et al., 2013), supporting the AP contention that infants have more than jaw-only control of articulators. However, there is uncertainty about the efficacy of such modeling because of limited articulatory data from infants (Nam et al., 2013).

To date, AP and FC accounts of CV preferences in babbling are based on corpus data and simulations, without direct empirical evidence of oral gestural control at the onset of babbling. As Goldstein et al. (2006) note, “while it is true that the preferred CV syllables could be produced by moving only the jaw, there is no direct evidence that the infants are in fact producing them in that way. And there is some indirect evidence that they are not” (p. 238). In other words, there is no direct evidence that infant oral constrictions are jaw-only driven or that infants at the onset of babbling have some control of lip and tongue independent of the jaw. At present, there is no direct means of measuring tongue articulations in babbling infants. As noted, however, AP has a clear and testable prediction about gesture imitation involving the underlying control of constriction organs independent of the jaw, namely the organ hypothesis (Goldstein, 2003; Studdert-Kennedy & Goldstein, 2003). The current study exploits the organ hypothesis in AP, aiming to provide empirical evidence on oral motor control structures at the emergence of babbling.

In the present study, we investigated 8-month-old infants’ oral responses to audiovisual presentation of repeated non-speech oral gestures. Specifically, we applied a cross-model comparison to evaluate infants’ responses to oral gestures (lip smacks and tongue smacks) produced by an adult female model. In line with AP, we predicted that infants would respond significantly more frequently with bilabial gestures to bilabial gesture presentation (lip smacks) than to the presentation of tongue-tip gestures (tongue smacks). This is contrary to predictions based on FC, where the response pattern to either stimulus condition would be quasi-random, and driven by jaw excursions only. We also predicted that they would respond more frequently with tongue gestures to tongue smack presentation than to lip smack presentation. In other words, we expected infants to respond with the matching articulatory organ to the model presentation (lips or tongue), but not necessarily in the same form as the adult gestures (cf., Meltzoff & Moore, 1997). For example, infants could produce repetitions of lip smacks with either protruded or inverted upper and lower lip postures, or they could show vocal tract constrictions with or without phonation. The critical issue for the present purpose was whether infants respond with the matching constriction organ.

2 Method

2.1 Participants

The participants were 18 full-term infants (12 girls, 6 boys) who were 8 months of age ($M = 35$ weeks, range 30–38 weeks). The infants were recruited from a large urban area and came from
North American English speaking, middle class households. The majority of the parents had received university-level education. According to parent report (Ages and Stages Questionnaires, Bricker & Squires, 1999), all participants were healthy, typically developing infants with no known problems in vision or hearing. An additional 16 infants were excluded because of fussiness \((n = 7)\), finger sucking \((n = 1)\), atypical oral behaviors (extended tongue protrusions) during testing \((n = 1)\), failure to complete the second session \((n = 3)\) or non-responsiveness to the stimuli despite good attention \((n = 4)\). Infant babbling development, as assessed by the first author (a speech-language pathologist) in interviews with the caregiver (cf., Oller, Eilers, Neal, & Cobo-Lewis, 1998) before each test session, revealed that 14 infants were in the canonical babbling stage (i.e., producing clear consonant-vowel repetitions) at their first visit, three transitioned into canonical babbling between the first and second visit, and one infant had not reached the canonical babbling stage by the second laboratory visit. In addition, caregivers were asked whether they had observed infant bilabial smacking or tongue smacking in recent spontaneous productions. According to parents’ report, 14 infants spontaneously produced lip smacks \((n = 8)\), tongue smacks \((n = 3)\) or both types \((n = 3)\) prior to the first test session. Anecdotal information showed that these behaviors often occurred in parent–infant interaction. All procedures were approved by the institutional research ethics board.

### 2.2 Apparatus, stimuli and procedures

The infants’ oral responses were recorded with a night shot camcorder (Sony Handycam DCR-HC28) during their observation of dynamic presentations of a female model producing bilabial smacks (LIP condition) or tongue smacks (TONGUE condition) on a monitor (21 inch LCD, Viewsonic VA903mb). Infants were tested separately for both stimulus conditions (LIP versus TONGUE) in counterbalanced order. The presentation conditions were divided over two sessions, scheduled 1 week apart, because the infants, on average, exhibited limited attention to stimulus presentation (mean duration = 5.9 min, \(SD = 2.7\) in the lip presentation condition; mean duration = 5.9 min, \(SD = 2.0\) in the tongue presentation condition). In the LIP condition, the model produced bilabial gestures with a smack sound but with less labial rounding than for a typical “kissing” gesture (see supplemental Video S1). In the TONGUE condition, smacks were produced with a constriction and release of the tongue tip at the alveolar ridge (see supplemental Video S2).

The infant was seated on the caregiver’s lap approximately 75 cm from the presentation monitor in a quiet and dimly lit testing room. Audiovisual stimuli were presented as follows (see Figure 1). Prior to stimulus presentation, infants were familiarized with the model’s smiling face for 10 s, followed by a 10-s baseline presentation consisting of five rhythmic audio clicks and simultaneous presentation of the model’s smiling face. Subsequently, lip or tongue smacks were presented in 30-s trials of six sets of smack repetitions (three smacks per set) at an average sound level of 65 dB. Each trial was followed by presentation of the model’s smiling face for 10 s to allow infants additional time to respond. After each trial the screen went black for 5 s. Trials were repeated until the infant showed distress by fussing, crying or moving around.

### 2.3 Analyses

Two trained coders independently analyzed the video files for lip and tongue gestures using ANVIL (versions 4.7.7. and 5.0) video annotation software (Kipp, 2001). Both visual and auditory information were used to identify the target behaviors because subtle oral movements (yawns, swallow, coughing) could be misinterpreted if annotation was solely based on visual information. Because it was impossible to separate the stimulus signal and infant response recordings (infants often
produced oral gestures during stimulus presentation), coders were necessarily aware of the presentation conditions. The videos were viewed in real-time and one frame at a time. Portions of the video recordings where infants were crying, yawning, fussing, sucking on fingers or toys, coughing or moving their face out of camera range were excluded from analysis. This resulted in 20% of the recorded data being excluded in each condition (LIP, TONGUE).

Infants’ attention to the stimuli (six sets of three smacks per trial) was determined by a perceptual based analysis of eye gaze (at screen or away) during the presentation of sets of smacks. The coders viewed the video recordings in frame-by-frame mode to determine changes in gaze direction. Inter-rater agreement for eye gaze annotations was determined for a random sample of 26 trials across five infants. Segmentation (start/stop gaze) agreement (78.7%) and category (at screen or away) agreement for gaze (95.8%; Cohen’s kappa = .92) were determined using ANVIL (version 5.0) agreement analysis. Subsequently, an association analysis was performed in ANVIL (version 5.0) to determine correlations for token pairs (smacks and response time versus looking at screen or away). The association between gaze and stimulus presentation was strong ($\chi^2 = 150.1$, $p < .001$). The results showed that the infants observed at least one gesture (one smack) during 81% of the sets of three LIP gestures and during 82% of the sets of TONGUE gestures.

Coding reliability for oral gestures was calculated from a random sample of 19% of the videos. Inter-coder segmentation agreement was 79% for lip gestures and 68% for tongue gestures. Small inter-coder differences in gesture onset and offset lowered the segmentation agreement. Inter-coder category agreement (CV, smacks, silent gesture, consonants) was 82% for lip gestures and 94% for tongue gestures with Cohen’s kappa $\kappa = 0.8$ and 0.9, respectively.

At the onset of a bilabial lip gesture, the upper and lower lip are (a) closed in a neutral position, with both lips touching across their entire length with a relaxed jaw position or (b) open in a relaxed position when lips are not touching and lips and jaw are relaxed. From this onset position the lips and jaw move into a bilabial constriction, with the lips moving from a neutrally closed to a more tightly closed position or from an open to closed position. Simultaneously, the jaw moves in an upward direction. Subsequently, the bilabial constriction is released into a relaxed open position and the jaw returns to its neutral position with a downward movement. At the onset of a tongue gesture the tongue is in a resting position inside the mouth, with either a relaxed open mouth or a

---

**Figure 1.** Schematic representation of the experimental paradigm.
closed mouth position. From this onset position the tongue tip moves (a) in a forward thrust between the upper and lower lip, (b) toward the upper lip with open mouth position, (c) in an upward movement toward the alveolar ridge. The tongue reaches the constriction fully before returning to the initial neutral position at the bottom of the oral cavity.

Gestures may be perceived by adults as belonging to several categories: (a) consonant-vowel-like sequences (VCV, CV), (b) single consonants (e.g., aspirated plosives), (c) non-speech sounds such as smacks or raspberries or (d) full oral constriction without phonation or other sound production (silent gestures). A gesture may be produced as a single constriction action or as part of a repetition of oral movements. In multi-cyclic repetitive movements there is no pause between the maximum jaw opening of the constriction release of one gesture and the onset of constriction for the following gesture. Gestures were considered part of a repetition when the duration between successive gestures was equal to or less than 1 s, and no audible breath occurred between two gestures.

Our dependent variable was the number of lip and tongue gesture events (single gestures and repetitions of gestures). One infant showed good attention to stimulus presentation but, unlike other infants, perseverated on lip-gesture productions (lip smacks and one type of bilabial CV production). This behavior was observed before and after as well as during the test session. This infant, who showed a much more frequent lip gesture response in both conditions than other infants, without any tongue gesture production, was considered an outlier and was added to the exclusion group. In addition, four sequences of gestures that contained heterorganic movement sequences (lip and tongue gestures) were excluded from the analyses (two in each presentation condition across all participants). Thus, only homorganic sequences were included in the analyses.

3 Results

The overall analysis included 123 LIP trials ($M = 6.83$ per infant; range 4–15 trials) and 112 TONGUE trials ($M = 6.22$ per infant; range 4–10 trials). Figure 2 displays the number of lip- and tongue-gesture events per presentation condition. A chi-square analysis with Yates correction revealed a significant association between type of gesture presentation (LIP, TONGUE) and oral gesture responses (lip, tongue), $\chi^2 (1, N = 272) = 78.4, p < .001$. Infants responded with more lip-gesture events (79%) than tongue-gesture events (21%) to LIP presentations, and they responded with more tongue-gesture events (75.8%) than lip-gesture events (24.2%) to TONGUE presentations.

A Wilcoxon signed-rank test to investigate individual response patterns across presentation conditions revealed that infants produced significantly more tongue-gesture events to TONGUE presentation than to LIP presentation ($Z = −3.17, p = .001$), but the across-condition difference was not significant for lip-gesture events ($Z = −1.56, p = .12$). Note, however, that response data were collected across two separate test sessions, which may have influenced cross-condition comparisons. There was a marginally significant trend for infants to produce more gesture events in the second laboratory visit than in the first ($Z = −1.90, p = .058$). Within presentation conditions, infants responded differentially, with more lip than tongue-gesture events to LIP gestures ($Z = −2.88, p = .004$) and more tongue than lip-gesture events to TONGUE presentations ($Z = −3.33, p = .001$). See supplemental Videos S3 and S4 for examples of infant oral responses to lip and tongue gesture presentation.

In both presentation conditions infants showed single gestures (55.9% of all gesture events) as well as gesture repetitions (44.1% of all gesture events). Bilabial repetitions contained on average 4.27 gestures ($SD = 1.95$), and tongue repetitions 3.25 gestures ($SD = 1.08$). Overall, such repetitions varied in length from 2 to 15 gestures, with the majority in the 2–5 range (86%). Individual gestures were coded for gesture type. Gesture repetitions contained a sequence of gestures of the same type (e.g., repetition of three bilabial smacks), or a sequence of varied gesture types (e.g.,
tongue smack – da – silent tongue gesture). Note, however, that repetitions were always produced with the same organ (tongue, lips). The majority of repetitions contained gestures of the same type (74%), with 26% of varied types.

Figure 3 shows the number of various gesture types per LIP and TONGUE condition. Infants responded to lip and tongue smack presentations with a variety of gesture types, including lip smacks, consonant-vowel (CV) combinations, single consonants, raspberries and silent gestures. In both LIP and TONGUE conditions CV combinations constituted the largest response category (48% of all gestures), followed by smacks (20%) and silent gestures (20%). Thirteen infants produced silent gestures as part of varied gesture repetitions. In fact, 84.8% of all silent gestures were part of a gesture-repetition series.

4 Discussion

In the present experiment, we tested contrasting predictions of AP and FC, specifically, the presence or absence of independent gestural control of lips and tongue in babbling infants. Such independent gestural control would be shown in organ responses (lip and tongue gestures) by 8-month-old infants matching an adult model’s oral gesture presentation. The results indicated that infants responded to the presentation of lip smacks with more bilabial gestures than tongue gestures, and they responded to the presentation of tongue smacks with more tongue than bilabial gestures. Within a specific category of organ responses (lips or tongue), gesture repetitions were varied, including CV sounds, consonants, smacks and silent gestures. In fact, 26% of all gesture repetitions consisted of sequences of mixed gesture types, for example, two identical CV syllables followed by two lip smacks. We did not perform within-type analyses, but individual differences in gesture form were observed. For example, some infants produced lip smacks with inverted lip postures, while others showed a more protruded constriction, in line with divergent patterns of facial mimicry in infancy (Meltzoff & Moore, 1977, 1997).

Figure 2. Number of gesture events (lips, tongue) per presentation condition (LIP, TONGUE).
Overall, the results suggest that 8-month-old infants have the ability to identify and control different constrictor organs independently from the jaw. The findings are consistent with the predictions of AP rather than those based on FC. According to the organ hypothesis in AP, infants should be capable of matching a model’s oral gestures because they make use of vocal tract actions in producing and perceiving oral gestures (Goldstein & Fowler, 2003). The implication is that infants can identify distinct constrictor organs from audiovisual information, and they can generate goal-directed, between-organ constrictions. Comparable goal-directed constrictions do not follow from FC, which assumes that constrictor locations are random at the early stages of babbling, depending on the initial position of articulators during jaw closing, and that infants lack control over lips and tongue independent of jaw oscillations. The present results are inconsistent with jaw-only dominance in early babbling, suggesting instead that infants can synchronize constrictor-directed activity of the tongue and lips with jaw oscillation (Goldstein, 2003).

Our results do not address CV biases in babbling directly. Rather, they focus on one fundamental difference between AP and FC, providing evidence against the FC view that vocal tract constrictor around the onset of babbling is controlled by the jaw alone. The differential organ responses, however, are in agreement with a gestural account of speech development. According to AP, individual gestures are combined into larger and increasingly complex structures with development. AP also proposes there is a relationship between the stability or strength of gestural coupling and their combinatoriality, which is reflected in developmental patterns. For example, synchronous coordination of gestures is presumed to emerge spontaneously because this is the most stable mode of coupled oscillators. This would explain the preference for CV combinations in babbling (Goldstein et al., 2006). Computational studies showed how the coupled oscillatory action of gestures might also account for other patterns in speech development. For example, the earlier appearance of consonant clusters in word coda compared to word onset is explained by a stronger coupling of all consonants to the vowel in word-initial position than in word-final position. The stronger CV coupling in the onset makes it more difficult to learn CC coordination at
word onset (Nam, Goldstein, & Saltzman, 2009). It goes beyond the scope of this paper to include a discussion on the development of syllable and word structure from an AP account, but studies investigating early word production describe how error in child speech can be explained as either paradigmatic gestural errors (e.g., error of degree or amplitude of a gesture) or syntagmatic errors (e.g., errors of timing and relative phasing of gestures) (Goldstein, 2003; Studdert-Kennedy & Goodell, 1995). This coupled oscillator approach to patterns in speech development implies that infants have access to functional oral synergies. Thus far, no empirical evidence existed showing that babbling infants have access to such goal-directed synergies. Results from the current study suggest they do. However, AP makes no clear prediction about when these functional synergies might develop, and further research in this direction is needed. Another limitation of AP is that it lacks an explicit account of the mechanism by which the child is able to extract relevant phase information from the articulatory and acoustic patterns in adult speech (Nam et al., 2009). Current results support the hypothesis that infants perceive vocal tract actions or gestures, but further testing is required, for example, by auditory-only gesture presentation using a similar paradigm as in the current study.

AP and FC accounts concur on phylogenetic and perhaps even ontogenetic relations between oscillatory movements in babbling and feeding (e.g., chewing, sucking). They disagree, however, on the control of lips and tongue in such behaviors. According to AP, infants can create distinct organ constrictions because they use their lips and tongue organs somewhat independently in feeding behaviors, drawing upon ancient mammalian oral capacities for feeding (Studdert-Kennedy & Goldstein, 2003). Thus, AP assumes independent constriction of vocal tract organs, relating this to lip and tongue smacks. FC suggests that the mandibular cycle underlying the speech syllable may have evolved from the mandibular cyclic movement in ancient mammalian ingestive functions, such as chewing, sucking and licking (MacNeilage, 1998). It assumes that this mandibular cyclic movement controls babbling in modern infants, but also speculates on a phylogenetic intermediate stage of the mandibular cycle in communicative behaviors such as lip smacks, tongue smacks and teeth chatters, before it was paired with phonation for syllable production (MacNeilage, 1998). In principle, lip and tongue smacks could be random constrictions with mainly jaw-only control, but we found systematic rather than random occurrences of lip and tongue smacks in response to the video presentations in the current study.

To date, the exploration of relations between vocalization and feeding behaviors has been limited to kinematic comparisons (Moore & Ruark, 1996; Steeve, 2010; Steeve, Moore, Green, Reilly, & McMurtry, 2008) in older infants, even though such behaviors are likely to exhibit altered kinematic characteristics after differentiation. Although both AP and FC propose an intermediate phylogenetic stage of smacking behavior as organisms progressed toward speech, this behavior has received little attention. In the current study, smacks were elicited in human infants in a highly controlled context. Parent reports indicated, however, that 14 infants produced spontaneous lip smacks at home, often in parent–infant interactions. A study of silent jaw wags in babbling infants also revealed spontaneous lip smacks in the home environment (Meier, McGarvin, Zakia, & Willerman, 1997).

Infants in the present study seamlessly combined phonated gestures (CV), unphonated gestures (silent gestures) and smacks in oscillatory movement sequences. This finding is consistent with previous reports of non-phonated (silent) jaw wags in combination with canonical babbling (Harold & Barlow, 2013; Meier et al., 1997). There are suggestions that the timing of silent jaw wags matches the timing of phonated jaw oscillations, indicating that infants at the onset of babbling may not have full mastery of the coordination of jaw movements (or articulation) with phonation and respiration (Meier et al., 1997). Such non-phonated, rhythmic oral gestures are typically excluded from the study of babbling and speech development (Moore & Ruark, 1996; Steeve, 2010; Steeve et al., 2008) or they are defined as spontaneous facial movements without a
specific function (Green & Wilson, 2006). We contend that infants at the onset of babbling vary their phonated and non-phonated jaw movements in a cyclic sequence of movements. Systematic longitudinal observation of infant lip smacks in a more ecologically valid context (e.g., home environment) could provide important insights into the transition from non-phonated to phonated oral behaviors.

Our observation of infants’ effortless transition between silent gestures, smacks and phonated CVs is relevant to Oller’s (2011) critique of AP and FC. Oller (2011) objects to reliance on phonetic transcriptions of babbling because such transcriptions were designed to characterize mature speech. Oller (2011) also objects to the failure to consider pre-canonical vocalizations, which are not in CV form. Oral rhythmic behaviors at the onset of babbling, including the non-phonated gestures observed here, may be of particular developmental significance, but they are difficult to access from the usual audio recordings. An approach to infant babbling based on vocal tract gestures, as in the present study, offers opportunities for investigating phonated and non-phonated vocal behavior with the same measures. Unquestionably, combined audio and video recordings are necessary to capture the full range of oral oscillatory behaviors in preverbal infants.

In summary, we found that 8-month-old infants showed organ-specific responses to audiovisual presentation of two types of oral gestures. They were capable of identifying and controlling distinct organ constrictions, in line with AP perspectives on the underlying control structures in babbling. Our findings also indicate that infants readily transition from phonated gestures, non-phonated gestures and smacks of lips and tongue. The time course of infants’ acquisition of within-organ gestural control and their coordination of gestures in intra- and inter-syllabic sequences await further research. The use of a controlled task, video annotation methods and a gestural framework for infant babbling, as in the present study, provides a productive approach to the study of mechanisms underlying babbling and the transition to speech.

Acknowledgements
We thank Jessica Davenport for assistance with video annotations and Julie Impey for assistance with stimulus design.

Funding
This work was supported by the National Sciences and Engineering Research Council of Canada (458145) and in part by the Canada Research Chairs Program.

Supplemental material
The supplemental Videos (S1, S2, S3 & S4) are available at http://las.sagepub.com/supplemental.

References


