2013

Effects of Environmental Stimulation on Infant Vocalizations and Orofacial Dynamics at the Onset of Canonical Babbling

Meredith Poore Harold
University of Kansas

Steven M. Barlow
University of Nebraska-Lincoln, steven.barlow@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/specedfacpub
Part of the Special Education and Teaching Commons
Effects of Environmental Stimulation on Infant Vocalizations and Orofacial Dynamics at the Onset of Canonical Babbling

Meredith Poore Harold¹ and Steven M. Barlow¹,²,³,⁴

¹Department of Speech–Language–Hearing: Sciences and Disorders, 3001 Dole, 1000 Sunnyside Ave, University of Kansas, Lawrence, KS 66049, TL: (785) 864-0630

²Department of Neuroscience, 3001 Dole, 1000 Sunnyside Ave, University of Kansas, Lawrence, KS 66049, TL: (785) 864-0630

³Department of Human Biology, 3001 Dole, 1000 Sunnyside Ave, University of Kansas, Lawrence, KS 66049, TL: (785) 864-0630

⁴Department of Bioengineering, 3001 Dole, 1000 Sunnyside Ave, University of Kansas, Lawrence, KS 66049, TL: (785) 864-0630

Abstract

The vocalizations and jaw kinematics of 30 infants aged 6–8 months were recorded using a Motion Analysis System and audiovisual technologies. This study represents the first attempt to determine the effect of play environment on infants’ rate of vocalization and jaw movement. Four play conditions were compared: watching videos, social contingent reinforcement and vocal modeling with an adult, playing alone with small toys, and playing alone with large toys. The fewest vocalizations and spontaneous movement were observed when infants were watching videos or interacting with an adult. Infants vocalized most when playing with large toys. The small toys, which naturally elicited gross motor movement (e.g., waving, banging, shaking), educated fewer vocalizations. This study was also the first to quantify the kinematics of vocalized and non-vocalized jaw movements of 6–8 month-old infants. Jaw kinematics did not differentiate infants who produced canonical syllables from those who did not. All infants produced many jaw movements without vocalization. However, during vocalization, infants were unlikely to move their jaw. This contradicts current theories that infant protophonic vocalizations are jaw dominant. Results of the current study can inform socio-linguistic and kinematic theories of canonical babbling.

Keywords

canonical babbling; speech; environment; kinematics

© 2012 Elsevier Inc. All rights reserved.

Additional contact information:
Meredith Poore Harold, meredithpoore@gmail.com
Steven M. Barlow, smbarlow@ku.edu

Publisher’s Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
1. Introduction

In the 12 months preceding an infant’s first word, a substantial amount of speech development occurs. The stages of infants’ pre-speech vocalizations are well-documented and consistent among socioeconomic, cultural, and linguistic backgrounds (Kent, 1984; Nathani, Ertmer, & Stark, 2006; Oller, Eilers, Neal, & Schwartz, 1999; Oller, 2000; Stoel-Gammon, 1992). These stages include phonation, followed by primitive articulation or goosing, followed by expansion or marginal babbling, and then canonical babbling. The onset of canonical babbling is between 5 and 10 months (Eilers, Oller, Levine, Basinger, Lynch et al., 1993; Ejiri, 1998; Oller, 2000; Stark, 1980). Canonical syllables consist of at least one full vowel and one full consonant, with adult-like timing and rapid formant transitions. These syllables occur in isolation first, and then reduplicated sequences (e.g., /bibi/, /dada/). The last pre-speech stage is variegated babbling, followed by the infant’s first words.

Perhaps the most researched and often cited stage of pre-speech vocalizations is canonical babbling. This is likely because canonical syllables are distinctly speech-like and easy for an adult to recognize and mimic (Oller, Eilers, & Basinger, 2001; Oller et al., 1999). Infants who are canonical babbling suddenly sound like they are talking. Canonical babbling also predicts later speech–language performance. The canonical syllable-consonant inventory predicts early word-consonant inventory (Stoel-Gammon, 1985). Infants who are delayed in the onset of canonical babbling (after 10 months) tend to have a smaller consonant inventory in the first year (Jensen, Bøggild-Andersen, Schmidt, Ankerhus, & Hansen, 1988), speech delay (Oller, Eilers, Neal, & Cobo-Lewis, 1998; Stark, Ansel, & Bond, 1988), expressive language delay (Jensen et al., 1988; Rvachew, Slawinski, Williams, & Green, 1999; Stoel-Gammon, 1992), receptive language delay (Jensen et al., 1988; Oller et al., 1998; Stoel-Gammon, 1989), and reading disability (Stark et al., 1988). Unfortunately, for over half of infants with delayed onset of canonical babbling, there is no earlier predictor—no diagnosed cognitive or medical problem, no noticeable atypical behavior or circumstances—to indicate imminent delay (Oller et al., 1999).

Why infants begin to produce canonical syllables in the first place, and make such a seemingly sudden transition from immature marginal syllables to precisely-timed canonical syllables is unknown (Locke, 1996). Infant speech researchers hypothesize that there is likely a core skill or set of skills that sparks canonical babbling onset. Though this could presumably be a linguistic, social, auditory, sensorimotor, and/or motor catalyst, the most common hypothesis is that it is a change in oromotor skills or control that allows an infant to begin to produce canonical syllables (Oller et al., 1999; Davis & MacNeilage, 1995).

This study represents the first investigation of the oral kinematic parameters of vocalizations from infants younger than 9 months. By comparing non-canonical babbling (NB) to canonical babbling (CB) infants, we hoped to identify what oromotor skills infants may be lacking before the onset of canonical syllable production, or what skills may allow them to begin to babble. Canonical syllable production is predicted to be dependent on mature oromotor skills (e.g., Moore, 2004; Oller, 2000), jaw movement skills in particular (Davis & MacNeilage, 1995, MacNeilage, 1998). Thus, the kinematics of the jaw and rate of vocalization and movement patterns were examined.

In addition to the lack of understanding of the skills needed to produce canonical syllables, it is also not known why infants babble at all, at any given moment (Locke, 1996). Infant speech–language researchers often spend an hour or more attempting to elicit vocalizations from young infants, switching from activity to activity (Fagan, 2009; Steeve & Moore, 2009; Steeve & Price, 2010). Within that hour, the researcher may use any or all of the following techniques: solo play with various toys, playing with experimenter, playing with
mother, watching videos, use of contingent reinforcement and modeling paradigms by the experimenter or mother, among others (Fagan, 2009; Nathani & Stark, 1996; Oller, Eilers, Steffens, Lynch, & Urbano, 1994; Stark, Bernstein, & Demorest, 1993; Steeve & Moore, 2009; Steeve & Price, 2010). These activities, all generally referred to as play, are generally poorly described and the rate of vocalizations across techniques remains unreported. Pilot data from our lab, however, indicated substantial differences among different environmental circumstances and the likelihood for eliciting vocalizations from an infant. For example, at the age of onset of canonical babbling, infants produce a substantial amount of arm stereotypy (e.g., hand waving, hand banging) (Ejiri, 1998; Iverson, 2005; Iverson & Fagan, 2004; Locke, Bekken, McMinn-Larson, & Wein, 1995; Thelen, 1981). We found that when given a toy that affords arm waving (grippable, lightweight, and is not easily dropped), infants this age wave their arms almost incessantly (Iverson, 2005; Iverson & Fagan, 2004), and that this fast arm waving appeared to interfere with the rate of vocalization. Also, there exists some contradictory evidence in the literature about how social interaction impacts the likelihood of vocalization in infants this age. Compared to older and younger infants, 6-month-olds are less likely to mimic the vocalizations of adults (Jones, 2007) and do not tend to direct their vocalizations toward others (Fay, 1967; Furrow, 1984). Between 5 and 9 months, infants vocalize more when they are alone than when an adult is interacting with them (Delack, 1978; Jones & Moss, 1971; Locke, 2004), but this disappears by the time they are 12 months. Age-appropriate modeling and contingent reinforcement of vocalization provided by an adult, however, are both techniques which have been suggested as appropriate for teaching and eliciting vocalizations from infants and toddlers (American Speech–Language–Hearing Association, 2008; Rosetti, 2001). This clash in evidence, along with other observations from pilot data preceding this study, demanded that the environmental circumstances during which infants vocalize required careful consideration and measurement.

2. Methods

2.1 Participants

This study included 30 infant participants 24–35 weeks old (6-month-olds: 11; 7-montholds: 8; 8-month-olds = 11). There were 12 females and 18 males (6-month-olds: 8 males, 3 females; 7-month-olds: 3 males, 5 females; 8-month-olds: 7 males, 4 females). Infants were recruited via Lawrence, Kansas birth announcements. Recruitment procedures, informed consent, and data collection procedures and methods for the current study were approved by the University of Kansas Human Subjects Committee. Eligibility to participate was determined by confirming that the infant was: 24–35 weeks old (6–8 months), from a Midwest monolingual American English household, full-term at birth (> 37 wks GA), currently healthy, and with no history of major medical illness (including vision, hearing, physical, or neurological impairments) nor diagnosed cognitive or motor disorder. An Ages and Stages 6-month (24–28 weeks) or 8-month (29–35 weeks) Questionnaire (ASQ; Paul H. Brookes Publishing Co., Inc.) was also completed at the infant’s scheduled appointment. ASQ scores within normal ranges and weight, length, and head circumference within the 5th–95th percentile (Centers for Disease Control and Prevention Infant Growth Charts, 2001) were required for data inclusion. The investigator also asked the parents open-ended questions about their infant’s vocalizations in order to determine vocalization stage (e.g. “Describe the sounds your baby makes”; procedures consistent with Oller et al., 1998, 1999, 2001). Open-ended questions were followed by specific questions to confirm canonical babbling status. Specific questions included verbal examples of canonical babbling modeled by the experimenter. Canonical syllables were also described as those that can be easily written down with alphabetic letters or easily mimicked by an adult, because they sound like syllables of American English. No parents provided evidence after specific questions that
conflicted in any way with what they said during open-ended questioning. As found by Oller et al., (1998, 1999, 2001), parents were very good at providing evidence as to whether or not their infant had reached canonical babbling status. Only infants who produced canonical syllables daily were classified as “Canonical Babblers (CB)”; \( n = 21 \). Infants who had never produced canonical syllables or had produced only a few canonical syllables ever were classified as “Non-Babblers (NB)”; \( n = 9 \).

### 2.2 Testing procedures

Upon arrival at the laboratory for testing, the study methods and equipment were explained to the parent(s) and consent for participation was obtained. During this time, the tester(s) talked and played with the infant to familiarize him/her with the new adults. In preparation for infrared digital video capture of mandibular movements (Motion Analysis Corporation, Santa Rosa, CA, USA), low-mass reflective markers (4 mm diameter) were attached to the infant’s head (for spatial reference) and off-midline mandibular edge (Figure 1, left). A Bolt-Beranek-Newman (BBN) miniature accelerometer was positioned on the infant’s skin over the thyroid lamina of the larynx (Figure 1, right). The accelerometer allowed for measurement of reactive tissue force in the area of the right thyroid lamina associated with vocal fold medialization and fundamental frequency oscillation during vocalization, grunts, and cough. The accelerometer provides a clean marker of laryngeal activity that is free of the usual acoustic artifacts due to environmental sounds (e.g., other talkers, toys banging on table surface), and was used in conjunction with an acoustic signal to identify vocalization onset and offset.

The infant sat in a modified high chair with specially designed 24” × 18” worktray, positioned near one wall of an 11’ × 12’ recording suite. Within the recording suite, the infant could see only the highchair s/he was sitting in, a blank black Samsung 40” LED monitor screen, and the illuminated diode array faces of five motion capture cameras, which were partially disguised with black covers. The parent sat in the motion capture suite, off to the right and behind the infant (3 feet), in order to help the infant feel safe and secure. The parent was instructed to remain motionless and quiet, not talk to or make eye contact with their infant, nor touch the infant or toys. The infant could see the parent with a 45° head turn, but very few infants ever looked at the parent during the 16 minutes of data collection.

### 2.3 Test session and conditions

Each infant participated in four successive four-minute test conditions, with condition order randomized, for a total of 16 minutes of data collection. The test conditions were as follows:

**LargePlay**—The infant was provided seven toys that were difficult to grasp with one hand (grip radius ranged from 9 cm to 34 cm circumference). Four were sound-producing (e.g., squeaky toy), three noiseless. The toys could be picked up and manipulated by the infant using both hands, but did not afford rapid play movements such as waving or shaking. All toys were presented at the same time. If the infant pushed a toy off his/her tray, it was ignored.

**SmallPlay**—Differed from LargePlay *only* in that the toys were easily grasped with one hand at the minimum circumference (grip radius ranged from 0.5 cm to 1.5 cm) and thus enabled the infant to produce rapid play movements (e.g., waving, banging, shaking). Four were soundproducing (e.g., squeaky toy), three noiseless; all toys were presented at the same time.

**Social**—The infant was not provided any toys on his/her tray. Instead, the experimenter sat next to the infant, brought each toy out one at a time, and used a strict sequence of modeling
and contingent reinforcement of vocalizations in order to attempt to elicit vocalizations from the infant. The social condition contained four activities (duck puppet, frog puppet, telephone, peek-a-boo) in counterbalanced order. (See Appendix A for condition script)

**Video**—The infant was not provided any toys to touch. Instead, s/he viewed a series of eighteen 13-second video clips of inanimate objects (e.g., spinning top, bubbles, moving train) cut from Baby Einstein™ videos. Thirteen-second clips with music and sound effects (~ 55 dB SPL) were alternated with seven-second silent video clips and presented in counterbalanced order. No clips were repeated, and sounded and silent clips were similar in nature, with the exception of the sound track.

### 2.4 Data acquisition and analysis

Kinematic data was collected using a video motion capture system (Motion Analysis Corporation, Santa Rosa, CA, USA) that included five Eagle-4.0 Mpixel digital real-time infrared cameras, programmed to sample at 119.88 frames/second, shutter speed at 1000/sec, and tracking parameters set with maximum speed to 30 mm/frame. Real-time data acquisition, synchronization, and display of all motion capture and biological signals were accomplished with Cortex™ software v.2.0.2.917 (Motion Analysis Corporation, Santa Rosa, CA, USA) running on a Quad-core MS WIN7 x64 bit workstation with dedicated 256 GB solid state drives to achieve maximum data throughput. The five two-dimensional (2D) camera views were co registered in real time by Cortex™ into a three-dimensional (3D) space by an acquisition routine. Reflective markers on the infant’s face were tracked with 0.15 mm resolution. Two acoustic data streams, laryngeal accelerometry, and an SVGA video data stream were digitized simultaneously by a National Instruments USB 6218 DAC (16-bit) multifunction I/O USB interface and merged synchronously by Cortex™ with the motion capture data channels. One acoustic data stream was sampled via a Sony cardioid directional microphone suspended six inches above the infant’s head, and used to align kinematic data within Cortex. The second acoustic channel was sampled via a Sony condenser microphone attached to the infant’s chair, approximately 6 inches from the mouth at shoulder level. Kinematic, acoustic, and accelerometry signals were digitized at 12 kHz/ channel at 16-bit vertical resolution (+/− 2.5V). Cortex™ was also used to analyze the video, acoustic, accelerometer, and kinematic data streams, along with a custom coded peak-picking software (MatLAB v.9 and LabVIEW), which helped identify relevant jaw movements. A jaw cycle (movement peak) was defined as any displacement of at least 2 mm plus any adjacent return toward baseline position within 1 second or less, from the initiation to termination of displacement. Jaw cycles were not required to start from a closed lip position, but could also start from a partially-open lip position.

Cycles were then categorized as either vocalized or silent, and with or without oral obstruction (e.g. finger or toy in mouth). Once the vocalized jaw cycles were identified, they were separated into one of three categories: non-protophonic (e.g. cry, squeal), vowel/marginal, or canonical. Protophonic vocalizations (vowel/marginal or canonical) that were not associated with a jaw cycle were also identified and counted. Previous research has shown that adult listeners are highly skilled at identifying canonical vs. non-canonical syllables (Oller, 2001). Inter-rater reliability for identification of vocalizations into non-protophonic, vowel/marginal, and canonical categories for the current study was 93%. All jaw cycles (silent [n = 1482], non-protophonic [n = 48], vowel/marginal [n = 97], and canonical [n = 43 ]) were then analyzed for the following parameters: jaw closing velocity (mm/sec), jaw opening velocity (mm/sec), and jaw cycle frequency (cps; different from jaw cycles per minute in that it is the timing parameter of an individual cycle). Examining jaw timing parameters was essential, because the most-often cited difference between canonical and marginal syllables is the change in acoustic timing parameters, and targeted vocal tract
resonances must be achieved by changes in kinematic spatiotemporal parameters (Oller, 2000). Other analyses included protophonic vocalizations per minute, jaw cycles per minute, and protophonic jaw cycles per minute. Of the 30 infant participants, 22 infants completed all four conditions. Reduced data sets were due to noncompliance during one or more conditions (e.g. crying), marker obstruction, or marker removal.

2.5 Statistical analysis

A mixed model statistical analysis was used in order to best account for the hierarchical nature of this data set (observations of outcome variables were repeatedly measured under different conditions [level-1], which were nested within subjects [level-2] and groups CB vs. NB [level-3]). Thus, we could do a bottom–up analysis where babbling status [level-3] was first accounted for, followed by the subject ID, age, and sex [level-2], followed by condition (LargePlay, SmallPlay, Social, Video) [level-3], and finally jaw kinematics. The mixed model analysis allowed for lack of independence among observations (data points for an individual infant as he she participated across a 16-minute time period) and missing data. The compound symmetry (CS) covariance structure of the repeated measures yielded smaller Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) than did the unstructured (UN), first-order autoregressive (AR), and variance component (VC) covariance structures, and thus was chosen for the current mixed model. For parameter estimation, maximum likelihood method was used. Statistical significance of model parameters was determined at 0.05 alpha level. All analyses were conducted using SAS 9.2 (SAS Institute, 2002–2008). Infants’ age (grouped by month), sex, and oral obstruction were included as covariates. When the group or condition effect was significant, adjusted means were pair-wise compared using a Bonferroni-corrected p-value.

3. Results

Significant effects were observed for the following CONDITION results:

- number of protophonic vocalizations produced in each condition \( F(3,83) = 183.72, p < .0001 \), Cohen’s \( f = 2.51 \) (large), \( p < .0001 \) for each comparison
- number of jaw cycles produced in each condition \( F(3,75) = 265.74, p < .0001 \), Cohen’s \( f = 3.17 \) (large), \( p < .0001 \) for every comparison except Social compared to Video
- number of protophonic jaw cycles produced in each condition \( F(3,72) = 85.35, p < .0001 \), Cohen’s \( f = 1.82 \) (large), \( p < .0001 \) for every comparison except Social compared to each SmallPlay and Video
- jaw opening velocity among conditions \( F(3,70) = 3.82, p = .01 \), \( p = .01 \) for only SmallPlay compared to Video with all other comparisons \( p > .30 \)
- frequency (cps) of jaw cycles produced in each condition \( F(3,70) = 10.42, p < .0001 \), Cohen’s \( f = .62 \) (medium), \( p < .0001 \) for LargePlay compared to SmallPlay and SocialPlay compared to Video, with all other comparisons \( p > .10 \).

No significant effect was observed for jaw closing velocity among conditions \( F(3,70) = 1.98, p = .13 \). Condition trends are summarized in Table 1 with the estimated mean and standard errors.

No significant effects were observed between NB and CB GROUPS and the: number of protophonic vocalizations produced \( F(1,26) = .10, p = .76 \); number of jaw cycles produced \( F(1,24) = .03, p = .86 \); number of protophonic jaw cycles produced \( F(1,24) = .32, p = .58 \); closing velocity of jaw cycles produced \( F(1,24) = .14, p = .71 \); opening velocity of
jaw cycles produced \( F(1,24) = .44, p = .52 \); or frequency of jaw cycles produced \( F(1,24) = .04, p = .84 \).

The following COVARIATE results are summarized. A significant effect was observed in infant sex and the number of protophonic vocalizations produced \( F(1,26) = 5.19, p < .05 \), Cohen’s \( f = .39 \) (medium) (Table 2). Females produced more protophonic vocalizations than males. This higher rate of vocalization in female infants has been observed in previous studies (Lewis, 1969; Lewis & Freele, 1973). No significant trends were observed in infant sex for number of jaw cycles produced, number of protophonic jaw cycles produced, jaw opening velocity, jaw closing velocity, or jaw cycle frequency.

No significant effects were observed in infant age across any measured parameters, including number of jaw cycles produced, number of protophonic vocalizations produced, number of protophonic jaw cycles produced, jaw opening velocity, jaw closing velocity, or jaw cycle frequency.

Oral obstruction required attention as a covariate because of the large amount of oral obstruction in many of the conditions. For each of the conditions, the percentage of cycles with oral obstruction was: LargePlay, 22%; SmallPlay, 27%; Social, 16%; Video, 10%. There was a significant effect observed in oral obstruction and jaw closing velocity \( F(1,24) = 53.16, p < .0001 \), Cohen’s \( f = 1.42 \) (large), and opening velocity \( F(1,24) = 27.02, p < .0001 \), Cohen’s \( f = .34 \) (medium). There was not a significant effect observed in oral obstruction and jaw cycle frequency \( F(1,24) = 6.07, p = .02 \). Oral obstruction trends are summarized in Table 3 with the estimated mean and standard errors. Because the frequency across jaw cycles with and without oral obstruction remains nearly constant, it is presumed that the increased jaw amplitude with intraoral obstruction reflects kinematic adaptation through a rescaling of opening/closing velocity in order to maintain cycle frequency (Table 3).

4. Discussion

4.1 Environmental catalysts of emergent canonical syllables

Infants produced the fewest protophonic vocalizations, jaw cycles, and protophonic jaw cycles per minute while watching videos, and this effect was significant when compared to each LargePlay and SmallPlay. Compared to Social, the effect reached significance in the measure of protophonic vocalizations per minute, but did not reach significance for jaw cycles or protophonic jaw cycles per minute. In sum, video stimuli encourage silence and stillness in infants 6–8 months of age. Some scientists have noted this effect in passing (e.g., Steeve & Moore, 2009). Others have found that videos don’t work well for rewarding contingent motor responses (Karzon & Banerjee, 2010), and that infants don’t learn well from videos compared to live interaction (DeLoache, Chiong, Sherman, Islam, Vanderborght et al., 2010; Richert, Robb, Fender, & Wartella, 2010). Though digital media has the benefits of being a convenient and easily-controlled input to infants, how infants process and conceptualize it is unknown. Whether the effect of videos on infants is temporary or compounding is also unknown, but the experience is certainly common and frequent—40% of 3-month-olds and 90% of 24-month-olds regularly watch television and videos (Zimmerman, Christakis, & Meltzoff, 2007).

Infants produced significantly more protophonic vocalizations and jaw cycles during the Social condition compared to Video. However, infants produced significantly fewer protophonic vocalizations and jaw cycles, per minute during social play with an adult compared LargePlay or SmallPlay, and significantly fewer protophonic jaw cycles per minute compared LargePlay. Thus, having an adult play with a 6- to 8-month-old infant is
less stifling to vocal and oromotor output than videos, but still more immediately stifling than allowing the infant to play with large or small toys alone. Importantly, the Social condition in this study used both age-appropriate modeling and contingent reinforcement of vocalization, which are both suggested as appropriate for teaching and eliciting vocalizations from infants and toddlers (American Speech–Language–Hearing Association, 2008; Rosetti, 2001). Also, plenty of silence (over 60%) was incorporated into the social condition, during which the adult gave the infant a chance to vocalize.

We suggest that there are several reasons for the lack of social vocalization in infants 6–8 months of age. First, it is important to note that change is the hallmark of infancy. Other studies have demonstrated that around the age of onset of canonical babbling, infants do not produce high rates of vocalization in the presence of adults or in social situations (Jones & Moss, 1971; Locke, 2004; Lin & Green, 2009; Delack, 1978) and they are unlikely to directly mimic the vocalizations of adults (Jones, 2007) or direct their vocalizations toward others (Fay, 1967; Furrow, 1984). We speculate that this is because infants are so attentive to and mesmerized by adults’ faces, eye contact, and behaviors at this age (Krentz & Corina, 2008; Kuhl, 2007; Nelson, 2001), that it temporarily stifles the infant’s own speech output. It is important to note, however, that a temporary reduction in speech output in no way overrides the substantial beneficial effects of adult social–vocal interaction to infants’ overall language development (Kuhl, 2007; Tamis-LeMonda & Bornstein, 2002; Warren & Brady, 2007). At this stage of vocal development, infants may simply not view canonical syllables as having a communicative function. Though infants certainly know that crying and vocalizing can elicit adult attention, this may a secondary factor for motivating the infant to vocalize. The infant may instead use vocalization primarily for the function of play and self-stimulation, or may not have any intent behind his or her vocalizations at all. Though the communicative function is certainly present or at least emergent by 12 months, there is no reason to assume that this is the case for younger infants, and this is likely the source for misunderstanding that social interaction is a proper way to elicit vocalizations from young infants.

Infants are by far the most vocal and active when playing alone with toys. LargePlay and SmallPlay elicited significantly more protophonic vocalizations, jaw cycles, and protophonic jaw cycles per minute than either Social or Video conditions. However, not all toys elicited the same type and amount of movement and vocalization from the infant. LargePlay elicited significantly more protophonic vocalizations and protophonic jaw cycles per minute than SmallPlay. We speculate that this is because arm waving and gross motor movements are not conducive to simultaneous vocalization, which requires the engagement of respiratory and phonatory musculature and motor control systems.

These results are not consistent with a previous study indicating that in the month of onset of canonical babbling, canonical syllables are more likely to occur with arm stereotypy than without (Iverson & Fagan, 2004). Looking at canonical syllables specifically, however, there was no preference for production during the SmallPlay condition at all. In fact, the overwhelming majority of canonical syllables (84%) were produced during the LargePlay condition, and SmallPlay elicited the same number of canonical syllables as the Social condition (both 7%), with the Video condition eliciting the fewest (2%).

Though LargePlay elicited the most protophonic vocalizations, SmallPlay elicited significantly more jaw cycles. We hypothesize that this increase in rate of jaw movements is due to the overall increased amount of bodily movement in this condition—increased large-amplitude and fast movements of the arms may carry over into control of the jaw and influence it to move a bit more. This increased activity level was also seen in other parameters—the SmallPlay condition elicited significantly greater jaw opening velocities.
(mm/sec) than the Video condition, and the frequency of jaw cycles was significantly higher in the SmallPlay condition compared to LargePlay and Video.

4.2 Orofacial kinematics and protophonic vocalizations of peri-canonical babblers

Comparisons between CB and NB infant groups were made in order to determine if there were differences between the oromotor skills of infants who were capable of producing canonical syllables (CB) versus those who were not (NB). We found no significant difference in the number of jaw cycles, protophonic vocalizations, or protophonic jaw cycles produced per minute between either NB and CB infants or among the tested age ranges of 6, 7, and 8 months. This result is consistent with data from a recent study (Fagan, 2009) indicating that infants do not demonstrate an increase in the number of syllables per utterance over the first year of life. However, it is inconsistent with the prediction that increases in jaw movement may hone infants’ oral kinematic skills and prepare him/her for the onset of canonical syllable production (Green & Wilson, 2006). Canonical babbling is thus marked by an increase in complexity, but not quantity of vocalizations or movements.

Researchers hypothesize that the onset in canonical syllable production requires some maturation of oromotor (Moore, 2004; Oller, 2000) and mandibular control in particular (Davis & MacNeilage, 1995; MacNeilage, 1998). Canonical syllables are distinct from marginal in that they have rapid formant transitions and adult-like timing of less than 500 milliseconds per syllable (Lynch, Oller, Steffens, Levine, Basinger, & Umbel, 1995; Oller, 2000), and this must be achieved by some oral kinematic adjustments. Because jaw speed and spatiotemporal coordination develop sooner than lips or tongue (Green, Moore & Reilly, 2002; Nip, Green, & Marx, 2009; Steeve, 2010) and vocalizations of infants from 9–24 months of age tend to be jawdominant (Green & Nip, 2010), the prediction for this study was that oral kinematic differences between NB and CB infants would be extractable from the jaw. Higher performance on speech–language tests among children 9 to 21 months of age has also been found to be associated with higher jaw speed (Nip, Green, & Marx, 2011).

In the current study, measured jaw kinematics were within the expected range, based on previous studies of older infants (Nip, et al., 2009) and studies of silent jaw motion in infants (Green & Wilson, 2006). However, we found no differences in jaw opening velocity, jaw closing velocity, or jaw cycle frequency across infants’ age or canonical babbling status. We suggest that the lack of jaw kinematic differences between babbling groups and across ages could be due to several factors. First, age was likely not a significant predictor simply because the age range of this study was too narrow to see a developmental trend. Other studies with wider age ranges (e.g., Nip et al., 2009) find significant differences in oromotor kinematics. Another possibility is that infants are either quite variable or at a semi-plateau in their development of jaw velocity and cycle frequency at this age.

The more surprising result was that there were no jaw kinematic differences between CB and NB infants. It may be that, at the time of transition to canonical syllables, infants are still learning this new kinematic skill and thus have quite variable motor performance. Dynamic Systems Theory suggests that the onset of a new level of skill is preceded by an increase in variability of motor performance (Smith & Thelen, 2003; Thelen, 1995), and both accurate and inaccurate attempts help refine the spatiotemporal parameters of the behavior (Smith, 2006). Many speech scientists suggest that speech motor development is nonlinear and marked by substantial amounts of variability, particularly at times of transition to attainment of a new skill (Goffman, 2010; Green & Nip, 2010; Smith & Zelaznik, 2004; van Lieshout, 2004). The inaccurate, possibly slower-velocity and slower-frequency jaw behaviors infants produce when newly canonical babbling may obscure the observation of faster and more accurate jaw movements.
Another reason for the lack of significant differences in jaw velocity and cycle frequency between NB and CB babies may be that overall increases in jaw velocity and jaw cycle frequency could actually occur *once* an infant begins to produce canonical syllables, rather than *in order for* the infant to begin to produce canonical syllables. The tasks demands of producing canonical syllables may drive neurophysiological maturation more so than neurophysiological maturation drives the onset of this new behavior.

A final reason for lack of oral kinematic differences between NB and CB infants may be due to analyses being performed across all the infants’ jaw movement types—silent, nonprotophonic, protophonic, and canonical. Oral kinematics may differ only for syllable type (e.g., marginal versus canonical), rather than across all jaw cycles and vocalization types. Unfortunately, much more data would need collected with either more infants or more trips per infant to the laboratory, because infants simply do not produce an adequate number of mature vocalizations in one data collection session in order to make comparisons. For example, there were 43 canonical syllables produced in this study. These syllables were from only 3 infants (thus, 18 babies who typically produce canonical syllables did not produce them during their lab visit), and 35 of these canonical syllables were from one infant.

A couple of interesting findings from the infants’ jaw kinematic data were the product of further exploratory data analyses. When comparing jaw cycles that had oral obstruction (e.g., finger or toy in mouth) to those that did not, the cycle frequencies of these two movement categories were nearly equal. Both opening and closing jaw velocity for cycles with oral obstruction, however, were substantially higher. Thus, infants seek to maintain constant temporal characteristics per movement episode, and timing has priority over movement speed. Infants are indeed capable of adjusting their jaw velocities, then, because they do so in order to account for the increased jaw amplitude caused by oral obstruction. This hierarchy of preserving temporal characteristics of jaw cycles and protophones will continue to be important later in development, as linguistic timing parameters will take precedent over articulatory speed when the oral cavity encounters spatial changes. The slower jaw velocities of younger infants compared to older infants (Nip et al., 2009) may be more linguistically-driven, instead of due to simple motor performance constraints.

A metric that did not reach significance when comparing across group or age, but may be noteworthy for future investigations was the measure of protophonic jaw cycles (protophonic vocalization with jaw movement). Of the total 140 protophonic jaw cycles produced, only 10 (7%) came from NB babies. Yet, there was no difference in either number of protophonic vocalizations nor number of jaw cycles per group, indicating that protophonic vocalizations with jaw displacement is a behavior produced primarily by CB infants. Though the great majority of protophonic jaw cycles were from CB babies, this measure was not significant in part because these protophonic jaw cycles were produced by only five of the total 21 CB infants. These five infants were all 8 months of age. We suggest that these five 8-month-old babies who do produce vocalization with jaw movement are demonstrating attainment of an advanced oral kinematic skill that neither group nor age adequately predict (see Figure 3).

Only 140 of the 810 protophonic vocalizations captured by the Motion Analysis System were accompanied by jaw displacement of ≥2 mm and cycle frequency ≥1 second. Thus, only 17% of 6–8-month-old infants’ vowels, marginal syllables, and canonical syllables include any jaw movement at all. The remaining 879 jaw cycles produced by the infants did not include vocalization. Thus, pairing the two together—vocalization plus jaw movement—may be an advanced pre-speech skill that is on the cusp of attainment as babies begin producing more and more canonical syllables. Previous theories of how infants coordinate...
jaw movement and vocalization, such as Frame–Content Theory (Davis & MacNeilage, 1995) emphasize that infant vocalizations are characterized by jaw movement plus vocalization. The results of this study indicate, however, that this type of movement is neither obligatory nor even the typical mode employed by infants 6–8 months of age (see Figure 3).

5. Conclusion

This study represents the first attempt to measure how the play environment affects infants’ rate of protophonic vocalization and jaw kinematics. The finding that changes in vocalization rate and kinematic parameters change with the presence of digital media, interactive adults, and certain types of toys is noteworthy for both researchers and clinicians (e.g., Speech–Language Pathologists) who seek to elicit vocalizations from infants in the most productive and efficient manner possible. This study was also the first to examine both protophonic vocalizations and jaw kinematics in infants ages 6–8 months, while the current literature base is focused almost exclusively on infants 9 months and older (e.g., Green & Nip, 2010; Nip et al., 2009). The findings that infants are able to modulate their orofacial speeds when presented with spatial changes, and that protophonic vocalizations at this age do not require jaw movement, as suggested by Frame–Content Theory, are noteworthy. Results from this study can inform theories of infant oral kinematic development for speech, as well as theories of infant social, cognitive, and linguistic development.

Acknowledgments

This work was supported in part by the Sutherland Family Foundation and the National Institutes of Health (R01 DC003311, Barlow-PI). Thanks to Dr. Jordan R. Green and Joan Wang, MSEE for technical, programming, and processing support. Special gratitude to Dr. Jaehoon Lee for statistical analysis.

REFERENCES


Iverson, JM. Increased manual rhythmicity precedes babble onset; Poster presented at the Biennial Meetings of the Society for Research in Child Development; Atlanta, GA. 2005.


Infant Behav Dev. Author manuscript; available in PMC 2014 February 01.
Highlights

- Jaw motion (Frame–Content Theory) does not accompany young infants’ vocalizations.
- Canonical babbling onset is not marked by changes in measured jaw kinematics.
- Modeling and contingent reinforcement do not elicit vocalization from young infants.
- Play environment affects vocalization and jaw movement in 6–8-month-old infants.
- Female infants (6–8 months) are nearly twice as likely to vocalize as males.
Figure 1.
Head block array and jaw marker (left); Jaw marker and laryngeal accelerometer (right)
Figure 2.
Test Conditions: LargePlay, SmallPlay, Social
Figure 3.
Protophonic jaw cycles by age, blocked into 6, 7, and 8-month-olds. Black circles = CB; White circles = NB
Table 1

Results by condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Protophonic vocalizations / minute</th>
<th>Jaw cycles / minute</th>
<th>Protophonic jaw cycles / minute</th>
<th>Jaw closing velocity (mm/sec)</th>
<th>Jaw opening velocity (mm/sec)</th>
<th>Jaw cycle frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SE</td>
<td>M</td>
<td>SE</td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>Large play</td>
<td>4.76</td>
<td>0.63</td>
<td>9.19</td>
<td>32.14</td>
<td>1.47</td>
<td>0.23</td>
</tr>
<tr>
<td>Small play</td>
<td>3.00</td>
<td>0.63</td>
<td>10.68</td>
<td>32.71</td>
<td>0.52</td>
<td>0.23</td>
</tr>
<tr>
<td>Social</td>
<td>1.79</td>
<td>0.64</td>
<td>2.72</td>
<td>34.13</td>
<td>0.19</td>
<td>0.24</td>
</tr>
<tr>
<td>Video</td>
<td>0.73</td>
<td>0.64</td>
<td>2.14</td>
<td>29.56</td>
<td>0.00</td>
<td>0.24</td>
</tr>
</tbody>
</table>
Table 2

Protophonic vocalizations per minute by sex

<table>
<thead>
<tr>
<th>Sex</th>
<th>M</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>2.86</td>
<td>.86</td>
</tr>
<tr>
<td>Males</td>
<td>1.55</td>
<td>.34</td>
</tr>
</tbody>
</table>
Table 3

Oral obstruction and jaw cycle kinematics

<table>
<thead>
<tr>
<th></th>
<th>Closing velocity (mm/sec)</th>
<th>Opening velocity (mm/sec)</th>
<th>Cycle Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>No oral obstruction</td>
<td>1311</td>
<td>27.06</td>
<td>0.55</td>
</tr>
<tr>
<td>Oral obstruction</td>
<td>359</td>
<td>38.89</td>
<td>1.46</td>
</tr>
</tbody>
</table>