The Emergence of Stability in Infant Vocal Production: A Model Based on Acoustic Analysis

Christopher Cox[1, 2], *Marilyn May Vihman*[2, 3]
[1] *Department of Linguistics & Cognitive Science, Aarhus University, Denmark*
[2] *Department of Language and Linguistic Science, University of York, UK*
[3] *University of California, Berkeley, USA*
chris.mm.cox@gmail.com, m.vihman@berkeley.edu

Abstract

Repeated babble production during early vocal development lays the groundwork for incipient motor control and creates stable motor-sensory mappings readily available for speech production. Investigation of this emergence of motor stabilisation has so far relied primarily on phonetic transcription and has therefore been unable to quantify continuous variability in speech production. In this longitudinal study, we leverage acoustic analysis to investigate how spectral centroids in the release phase of 5634 plosives change in 28 infants between the ages of 9 and 18 months. Modelling developmental change in hierarchical Bayesian location-scale models, we find a reliable non-linear increase followed by a gradual decrease in production variability. We argue that this U-shaped developmental trajectory reflects a point of transition, with exploration and subsequent consolidation of new vocal behaviours.

Introduction

Babble is a rhythmic vocal behaviour that emerges at around the same time as rhythmic stereotypies in other motor areas (Kent, 1984; Thelen, 1981) in both humans and non-human animals (Davis & MacNeilage, 2000; Lindblom, 2000; Leitão & Gahr, 2024), regardless of ambient language or amount of input (Peute & Casillas, 2022). Between the emergence of consonant-vowel syllables in babble (6-8 months: Oller, 2000) and first word production (typically about 3-5 months later: Fagan, 2009), infants' vocalisations tend to settle on two or more well-practised vocal routines that remain frequent and stable across longitudinal observations (McCune & Vihman, 2001). The emergence of these 'vocal motor schemes' constitutes a pivotal moment in vocal development because it allows infants to discover and explore sensory-motor associations (Vihman et al., 2014) and supports retention of word forms (Keren-Portnoy et al., 2010; Vihman, 2022). Experimental and observational work has shown that a child's individual production experience exerts an influence on their attention to sounds in the speech stream (DePaolis et al., 2011, 2013; Majorano et al., 2014; Laing & Bergelson, 2020; Vilain et al., 2019) and creates individual trajectories of downstream development in babble and word production (McCune & Vihman, 2001; Keren-Portnoy et al., 2010). Children thus play an active role in shaping their own vocal development (Elmlinger et al., 2019, 2023; Goldstein & Schwade, 2008; Laing & Bergelson, 2020). Longitudinal studies of continuous change in infants' production patterns hold the potential to improve our understanding of variability in these individual developmental trajectories.

Aims of the current study

The identification of the emergence of vocal motor schemes are typically based on phonetic transcriptions (McCune & Vihman, 2001). Children, however, often produce speech sounds that fall between traditional target sounds (Kent &

Murray, 1982; Frisch & Wright, 2002) and exhibit more protracted production variability when assessed by acoustic analysis than by transcription alone (Lee et al., 1999; Munson, 2004). In the following analyses, we leverage acoustic analysis to measure continuous change in children's speech production and explore the following questions:

i. What are the developmental trajectories of production variability at each place of articulation as the infant starts to stabilise their motor routines?
ii. How do these trajectories relate to transcription-based time points of vocal motor scheme attainment?

Material and methods

Participants

The speech recordings for this study consisted of vocalisations from 28 infants recorded every two weeks between the ages of 9 and 18 months. All vocalisations were transcribed, and each child was coded for vocal motor schemes from these transcribed records according to the following two criteria from previous studies (McCune & Vihman, 2001; DePaolis et al., 2011; Majorano et al., 2014). A consonant:

i. is produced at least 10 times in each of three out of four successive sessions, or
ii. occurs 50 times within one to three sessions.

These criteria attempt to capture the consistency and frequency that characterise the individual production experience of each child (McCune & Vihman, 2001). To observe how acoustic variability develops around the point of VMS attainment in this study, we converted the time point of vocal motor scheme attainment for each infant to a continuous predictor by computing the number of days from vocal motor scheme attainment for each child. We used days from *second* vocal motor scheme attainment in the following analyses, as this marks the onset of emergent vocal control and phonological systematicity (McCune & Vihman, 2001).

Acoustic analysis

The onset and offset of each burst in the 5634 plosives (1788 bilabial, 2709 coronal, 1137 dorsal) were manually marked with boundaries in Praat (Boersma and Weenink, 2013). Recordings were resampled at 16 kHz, pre-emphasized above 1000 Hz, and high-pass filtered at 200 Hz to reduce the influence of low-frequency glottal vibration (Forrest et al., 1988). For each burst, a smoothed spectrum was calculated by averaging the squared amplitude values of seven 64-point FFT spectra taken from 3ms Hamming windows. As our primary measure, we calculated the power-weighted average frequency of the spectrum (i.e., centre of gravity, or spectral centroid). The centroid at plosive release depends on the cavities surrounding the point of constriction; for example, we would expect dorsal plosives to exhibit a lower spectral centroid than coronal plosives because the cavity in front of the dorsal release is larger and has a lower resonant frequency (Chodroff & Wilson, 2014).

Statistical Models

Hierarchical Bayesian location-scale models were fitted to the acoustic data with weakly informative priors. To model how the location (centroid) and scale (variability) of the spectral centroids at each place of articulation change over developmental time for each participant, we allowed the intercepts and slopes for both the dependent variable and the sigma to vary for each participant. Because the three places of articulation (i.e., bilabial, coronal and dorsal) may exhibit different developmental trajectories, we estimated change in the sigma parameter for each of the three places of articulation and added

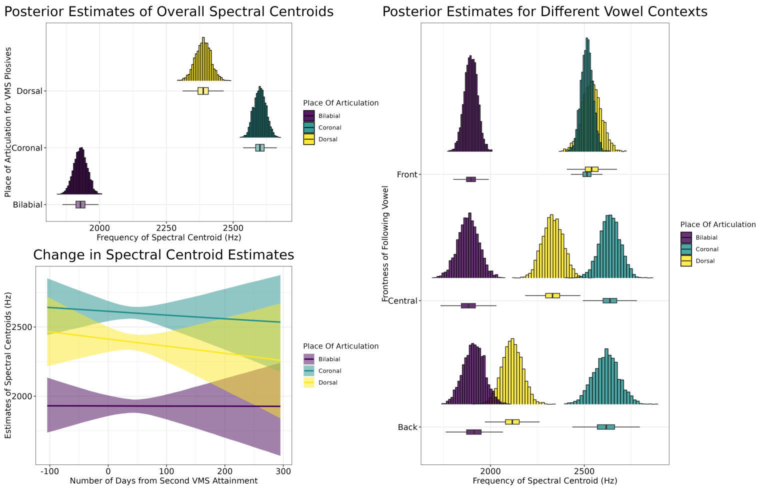

Figure 1: Panel of three plots of i) Posterior estimates of spectral centroids across bilabial, coronal, and dorsal plosives (top-left); ii) Estimates as a function of days from attainment of second vocal motor scheme (VMS) – 0 indicates the day each child attains their second VMS (bottom-left); iii) Posterior estimates for place of articulation of plosives before different vowel contexts (right).

a factor smooth spline term to capture potential non-linear changes in production variability.

All computations were performed in R 4.2.0 (R Core Team, 2020) using brms 2.16 (Bürkner, 2017) and cmdstanr 2.28.2 (Carpenter et al., 2017). We chose weakly informative priors in or-der to discount extreme values as un-likely and to ensure minimal influence on the posterior estimates of the model. In what follows, we provide estimates and report 95% credible intervals and evidence ratios. The credible interval refers to the range of values that have 95% probability of containing the true parameter value given the assumptions of the model. We report these intervals in square brackets. The evidence ratio denotes the ratio of likelihood in favour of a hypothesis. An evidence ratio of 10, then, implies that the hypothesis is 10 times more likely than the alternative. An evidence ratio of ‘Inf’ (infinite) occurs when all the posterior samples conform to the direction of the hypothesis and not to alternative directions.

Results

Spectral Centroids

The spectral centroid estimates exhibited differentiation according to place of articulation (bilabial: 1929.15 Hz [1880, 1977], coronal: 2603 Hz [2559, 2648], dorsal: 2384 Hz [2327, 2440]) (Figure 1, top-left) and showed no reliable changes with development (bilabial: -1.39 Hz [-122, 111] ER: 1.03; coronal: -26.26 Hz [-150, 101], ER: 1.91; dorsal: -48.15 Hz [-197, 103], ER: 2.76) (Figure 1, bottom-left). Fronter following vowels (Figure 1, right) made no changes to bilabial plosives (ER ≈ 1), but raised spectral centroids for dorsal plosives (335 Hz [232, 435], ER: Inf) and lowered them for coronal plosives (-102 Hz [-189, -15], ER: 33).

Variability in Spectral Centroids

Production variability in the plosives showed a non-linear decrease over time (Figure 2, left). Coronal plosives exhibited more variability (796 Hz [757, 846]) than bilabial (659 Hz [626, 699])

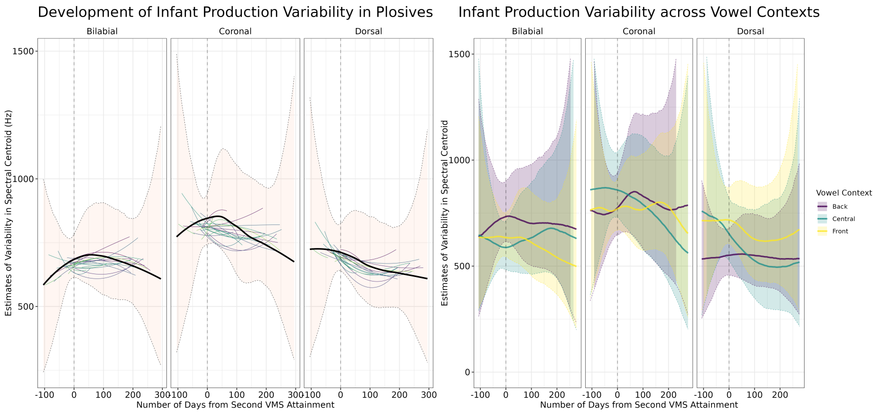

Figure 2: Panel of two plots of i) Posterior estimates of the development of sigma (i.e., variability) across bilabial, coronal, and dorsal plosives where coloured faded lines represent individual child trajectories and orange faded area show 95% credible intervals (left); ii) Posterior estimates of the development of sigma for plosives before different vowel contexts where faded areas show 95% credible intervals (right).

or dorsal (679 Hz [633, 721]) plosives (ER: Inf). The time point of peak production variability was on average within 27 days of attainment of two vocal motor schemes; however, these vertices exhibited a high degree of variability across children and place of articulation (SD = 69 days). Bilabial plosives exhibited more variability before back vowels (706 Hz [626, 796], ER: Inf), and dorsal plosives showed more pro-duction variability before front vowels (679 Hz [602, 765], ER: 442) (Figure 2, right).

Discussion

By leveraging acoustic analysis to measure continuous change in infants' production of plosives, we showed that the spectral centroids exhibited clear differentiation as a function of the tran-scribed place of articulation, no reliable developmental changes, and robust ef-fects of coarticulation with the follow-ing vowel (Figure 1). We also found evidence for non-linear patterns of de-velopment in variability across all three places of articulation; that is, an initial increase in production variability (espe-cially in bilabial and coronal plosives) followed by a gradual decrease over time (Figure 2).

These findings indicate that infants can draw upon their babble experience to succeed in executing crude articulatory production targets (cf., Kent, 2022; McCune & Vihman, 2001). However, the integration of these articulatory movements into dynamic motor constellations remains characterised by a high degree of production variability, the peak of which falls shortly after the transcription-based point of attainment of the second vocal motor scheme (McCune & Vihman, 2001). The finding of peak variability around the time point of infants' attainment of their second vocal motor scheme accords with studies showing that variability is characteristic of the plasticity of an open system (Studdert-Kennedy, 1986), and that a reduction in variability often co-occurs with the stabilisation of adap-tive motor behaviours (Piek, 2002; Thelen et al., 1987; Vihman, 1993). The tendency for infants to exhibit more production variability in bilabial plosives before back vowels and in dor-sal plosives before front vowels may indicate that infants are still learning to control the dynamics of articulatory transitions, from various types of con-strictions to more open configurations of the vocal tract (cf.

Lindblom, 2000). The above finding of a gradual decrease in production variability over time in parallel with no reliable change in the location of the spectral centroids indi-cates that with repeated practice and the gradual stabilisation of these sen-sorimotor routines over time, infants build up a repertoire of consistent pho-netic output forms (Vihman, 1993; Ekström, 2022). This repertoire of motor constellations in turn leads to the formation of phonetic concepts in memory and shapes infants' attention to the speech input through the 'articulato-ry filter', which leads them to retain word forms that contain the sounds they can most easily reproduce (Vi-hman, 2022).

Conclusion

By investigating continuous change in the variability of infants' motor routines, this study adds to our under-standing of how infants transition from the variable vocalisations they display early in life to the stable patterns need-ed for word production. The sudden increase and subsequent gradual de-crease in variability among the plosives speak in favor of a production system in a period of transition. During this early stage of vocal development, the infant is learning to connect the phonetic gestures involved in speech with their corresponding acoustic patterns. The child's repeated use of these motor routines leads to the stabilisation of phonetic concepts in memory and reduction of production variability over time.

References

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3. 51. *Online: http://www. praat. org*, *2*.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, *80*, 1-28.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*.

Chodroff, E., & Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *The Journal of the Acoustical Society of America*, *136*(5), 2762-2772.

DePaolis, R. A., Vihman, M. M., & Keren Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants?. *Infant Behavior and Development*, *34*(4), 590-601.

DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, *36*(4), 642-649.

Ekström, A. G. (2022). Motor constellation theory: A model of infants' phonological development. *Frontiers in Psychology*, *13*, 996894.

Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019). The ecology of prelinguistic vocal learning: Parents simplify the structure of their speech in response to babbling. *Journal of child language*, *46*(5), 998-1011.

Elmlinger, S. L., Schwade, J. A., Vollmer, L., & Goldstein, M. H. (2023). Learning how to learn from social feedback: The origins of early vocal development. *Developmental Science*, *26*(2), e13296.

de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, *67*(2), 297-319.

Fagan, M. K. (2009). Mean length of utterance before words and grammar: Longitudinal trends and developmental implications of infant vocalizations. *Journal of child language*, *36*(3), 495-527.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, *84*(1), 115-123.

Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*(2), 139-162.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological science*, *19*(5), 515-523.

Kent, R. D. (1984). Psychobiology of speech development: Coemergence of language and a movement system. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *246*(6), R888-R894.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, *72*(2), 353-365.

Kent, R. D. (2022). The maturational gradient of infant vocalizations: Developmental stages and functional modules. *Infant Behavior and Development*, *66*, 101682.

Keren-Portnoy, T., Vihman, M. M., De Paolis, R. A., Whitaker, C. J., & Williams, N. M. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language and Hearing Research 53*(5), 1280-1293

Laing, C., & Bergelson, E. (2020). From babble to words: Infants' early productions match words and objects in their environment, *Cognitive psychology*, *122*, 101308.

Leitão, A., & Gahr, M. (2024). Babbling opens the sensory phase for imitative vocal learning. *PNAS*, *121*(18), e2312323121.

Lee, S. A. S., & Iverson, G. K. (2012). Stop consonant productions of Korean–English bilingual children. *Bilingualism: Language and Cognition*, *15*(2), 275-287.

Lindblom, B. (2000). Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica*, *57*(2-4), 297-314.

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*(5465), 527-531.

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, *10*(2), 179-204.

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development. *Journal of Speech, Language and Hearing Research, 44*(3), 670-684.

Munson, B. (2004). Variability in /s/ production in children and adults. *Journal of Speech, Language and Hearing Research, 47*(1), 58-69.

Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.

Piek, J. P. (2002). The role of variability in early motor development. *Infant Behavior and Development*, *25*(4), 452-465.

Peute, B., & Casillas, M. (2022). (Non-) effects of linguistic environment on early stable consonant production: A cross cultural case study. *Glossa: a journal of general linguistics*, *7*(1).

Studdert-Kennedy, M. (1986) Sources of variability in early speech development. In Invariance and variability in speech processes (J. S. Perkell & D. H. Klatt, editors), pp. 58-76. Hillsdale, NJ: Lawrence Erlbaum.

Team, R. C. (2020). A language and environment for statistical computing. *R Foundation for Statistical Computing*.

Thelen, E. (1981). Rhythmical behavior in infancy: An ethological perspective. *Developmental psychology*, *17*(3), 237-257.

Thelen, E., Kelso, J. S., & Fogel, A. (1987). Self-organizing systems and infant motor development. *Developmental review*, *7*(1), 39-65.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, *21*(1-2), 61-82.

Vihman, M. M., DePaolis, R. A., & Keren Portnoy, T. (2014). The role of production in infant word learning. *Language Learning*, *64*(s2), 121-140.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language*, 397-445.

Vihman, M. M. (2022). The developmental origins of phonological memory. *Psychological review*, *129*(6), 1495-1508.

Vilain, A., Dole, M., Lœvenbruck, H., Pascalis, O., & Schwartz, J. L. (2019). The role of production abilities in the perception of consonant category in infants. *Developmental science*, *22*(6), e12830.