

UC Berkeley

UC Berkeley Previously Published Works

Title

Babble and First Word Learning: The Articulatory Filter Revisited

Permalink

<https://escholarship.org/uc/item/6d75r3gk>

Journal

Languages, 11(9)

ISSN

2226-471X

Authors

Vihman, Marilyn M

Cox, Christopher MM

Publication Date

2026-09-01

DOI

10.3390/languages11090187

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Review

Babble and First Word Learning: The Articulatory Filter Revisited

Marilyn M. Vihman ^{1,2,*}  and Christopher M. M. Cox ³ ¹ Department of Linguistics, University of California, Berkeley, CA 94704, USA² Department of Language and Linguistic Science, University of York, York YO10 5DD, UK³ Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, 8000 Aarhus, Denmark; chris.mm.cox@gmail.com

* Correspondence: m.vihman@berkeley.edu

Abstract

Production and perception of speech develop in tandem during the first year of life, shaping one another in ways that have only recently been experimentally tested. Canonical babbling, which emerges robustly between 6 and 8 months across languages, is grounded in neuromotor maturation but also depends critically on audition. To account for how production experience might shape word learning, in 1993 Vihman proposed an ‘articulatory filter’ that spotlights adult speech patterns that match an infant’s emerging motor routines. This review traces this proposal to its origins and summarizes the experimental evidence from the past 15 years, much of it using Vocal Motor Schemes (VMS)—frequent, stable consonantal patterns identified from naturalistic recordings—as an index of phonetic mastery. Studies employing VMS demonstrate that infants’ production patterns modulate their attention to and memory for both nonwords and familiar words, and that earlier VMS attainment predicts later lexical and phonological outcomes. Recent acoustic work has begun to validate the VMS construct, while parallel lines of research extend the logic to the pre-canonical period and to infants’ attraction to the sound of infant voices, especially their own. With further validation, VMS may serve as an early clinical indicator of vocabulary delay.

Keywords: articulatory filter; babble; vocal motor scheme; vocal production effects on speech perception

1. Babble and the Sensorimotor Foundations of Speech

How does an infant’s own vocal production shape the way the infant processes and remembers speech? This question, raised at the dawn of psycholinguistic research on child language (Fry, 1966), has gained empirical traction over the past two decades. The vocal landmark in focus—the reliable occurrence of canonical babble in the first year—was well established by the stage models of the 1980s, based on infants learning English (Oller, 1980; Stark, 1980), Dutch (Koopmans-van Beinum & Van der Stelt, 1986) and Swedish (Roug et al., 1989). To our knowledge, no evidence of ambient language differences in the timing or identity of the first canonical babble produced in different ambient language environments has been reported. Instead, its consistent emergence at about 6 to 8 months (Oller, 1980, 2000)¹ appears to reflect the maturation of central neuromotor programming, which provides a substrate for coordinated movements in general (Thelen, 1981).

The robust timing of canonical babble is consistent with its place in a broader class of rhythmic motor behaviors. Thelen’s ‘rhythmic stereotypies’ (kicking, waving, banging,



Academic Editors: Susan Rvachew
and Margaret Kehoe

Received: 28 May 2026

Revised: 26 August 2026

Accepted: 2 September 2026

Published: 15 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

rocking, etc.)—repetitive movements produced in sustained bouts with no apparent instrumental purpose—peak in frequency at about 6 to 7 months, overall, and canonical babble follows this timetable as well. More specifically, Ejiri (1998) established, in a longitudinal study of 28 Japanese infants, that the onset of rhythmic canonical babble corresponds with the peak of rhythmic hand actions, at around 6–7 months. The parallel between babble and rhythmic manual activity rests on this contingency finding, together with the evidence from deafness and cochlear implantation, discussed below, in which motor readiness and auditory access diverge.

Coinciding onsets of this kind (i.e., behaviors that emerge together in development) are ubiquitous in the first year, however, and are in themselves compatible with any number of causal arrangements. The case for a genuine link between babble and rhythmic manual activity rests on two further findings presented below: Ejiri’s observation that infants’ rattle-shaking comes to favor a sound-producing rattle only after babble onset, and the evidence from deafness and cochlear implantation, where the contributions of motor readiness and of hearing can be separately observed (on such motor-language ‘cascades’, in which advances in one developing system propagate to another, see Iverson, 2010, 2021). Thus, the robust timing and universal quality of these first adult-like vocalizations suggest that they are rooted in neuromotor maturation.

Audition plays a critical role both in supporting the emergence of canonical babble and in shaping infants’ responses to this new vocal activity once it is underway. That is, access to input speech is needed for its timely emergence (Oller & Eilers, 1988; Rvachew et al., 1999). Ejiri (1998) provides the interesting finding that in her experimental condition, although the rhythmic shaking of a rattle accompanied the emergence of canonical babble whether the rattle produced a sound or not, a relative increase in the rhythmic shaking of the sound-making rattle was observed only after canonical babble onset. This suggests that infants first move both hands and articulators rhythmically as those movements become available to them; the onset of adult-like syllabic production then arouses interest in the audible consequences of babble and, incidentally, of the shaking hand.

The role of self-generated auditory feedback in sustaining babble is well-exemplified by what is known of deaf and cochlear-implanted infants. Given that silent babble or ‘jaw wagging’ has been observed in deaf as well as hearing infants in the first year (Meier et al., 1997), the motorically grounded rhythmic stereotypy of canonical babble seems to be available to deaf infants as well. However, although profoundly deaf infants may produce some early canonical syllables, they typically do not go on to elaborate and sustain canonical babbling (Fagan, 2015).

Cochlear-implanted infants offer a natural experiment that decouples motor readiness from auditory access. Following cochlear implant activation, infants progress through the typical sequence of prelinguistic stages on a timeline indexed to auditory experience rather than chronological age, often reaching canonical and advanced babbling milestones with fewer months of hearing experience than typically developing infants require (Schauwers et al., 2004; Ertmer & Jung, 2012; for a systematic review of post-implantation growth trajectories and measurement heterogeneity, see McDaniel & Gifford, 2020). Reduplicated babbling, in particular, is almost entirely absent before implantation but reaches hearing-peer levels within roughly four months of activation, suggesting that extended production of reduplicated babble is specifically driven by infants’ motivation to generate the associated auditory feedback (Fagan, 2015). In other words, the anatomical and motor machinery for canonical babble appears to be largely in place before implantation, but its consolidation into sustained, exploratory production depends on infants being able to hear and act on the speech-like character of their own vocal output.

The sensorimotor link is key to the role of vocal production in word learning. Canonical babble can be taken to emerge as ‘unsupervised learning. . . in which stable action-perception links are acquired through random articulatory practice with auditory and somatosensory feedback’ (Rvachew & Alhaidary, 2018, p. 8). The sensorimotor link no doubt antedates babble, as seen in recent studies (discussed below). However, it is canonical babble that provides the infant, through automatic feedback, with the repeated experience of a self-produced semblance of adult speech. As Rvachew and Alhaidary point out, ‘self-supervised learning is . . . possible. . . [with] early learning . . . intrinsically motivated and focused on self-generated auditory targets’ (p. 9). This is essentially the vocal advance captured by the proposal detailed below.

Over 30 years ago, in response to the question, ‘How does the infant perceive and produce speech sounds?’ (Beckman, 1993, p. 1), Vihman introduced the concept of an ‘articulatory filter’, posited to relate the child’s small set of well-established vocal routines, or motor plans, to globally similar patterns in adult speech:

It is through the mechanism of attention to his or her own babbling, or vocal exploration, that the child discovers the link between the phonetic gestures underlying speech and the acoustic patterns that accompany them. . . This ‘discovery’ gives rise to what we may term an ‘articulatory filter’, a phonetic template (unique to each child) which renders similar patterns in adult speech unusually salient or memorable; in particular, the filter picks out patterns for which the child has already established a ‘motor plan’. . . specifying the articulatory implementation which will result in the particular sound pattern. (Vihman, 1993, p. 74)

This review traces the history of this concept back to the first suggestion of such an effect of production on word learning and then carries it forward to consider the evidence that has accumulated over the past 15 years. The review also includes a brief overview of more recent work proposing articulatory effects on speech processing in the pre-canonical period, and of studies proposing that infant attraction to the sound of infants’ voices—especially their own—may be a factor meriting more attention. The conclusion returns to the role of vocal production in infant development and word learning and the ways in which the idea of an articulatory filter might be further tested and applied.

2. Origins of the Articulatory Filter

Fry (1966) seems to have been the first to suggest that the emergence of syllable-based vocal production (or canonical babble) is critical for word learning. He offered two reasons for this. First, it provides the child with experience coordinating ‘the phonatory and articulatory muscle systems’ (p. 189). Specifically,

The child is ‘getting the idea’ of combining the action of the larynx with the movements of the articulators, of controlling to some extent the larynx frequency, of using the outgoing airstream to produce different kinds of articulation, and also the idea, which is quite important, of producing the same sound again by repeating the movements. (p. 189).

Second, vocal practice makes it possible to establish an auditory feedback loop.

As sound-producing movements are repeated and repeated, a strong link is forged between tactual and kinesthetic impressions and the auditory sensations that the child receives from his own utterances. (p. 189)

Citing Fry (1966), and noting also Locke’s (1986) suggestion that children at the early stage of word production ‘consult their store of available articulations in search of the closest matches’ (Vihman, 1991, p. 76), Vihman formulated the hypothesis that a child’s early word forms (typically, one to two syllables long, cross-linguistically: Vihman, 2019)

will be constrained by their existing motor practice: ‘The child will...attempt to produce an adult form only when it matches one of the child’s more common babbled vocalizations, or is similar enough to evoke such a form’ (Vihman, 1991, p. 77).

In 1996, still drawing only on production data, Vihman expressed the idea a little differently:

The child may be seen as experiencing the flow of adult speech through an ‘articulatory filter’ which selectively enhances motoric recall of phonetically accessible words...The earliest recognizable word productions would then be a product of the child’s experience of a match, in familiar situational context, between a commonly produced adult form and his or her own babble forms...The production filter on perception of phonetic patterns... is taken to be the product of the ongoing strengthening—due to the combined effects of proprioceptive and auditory feedback—of emergent child vocal patterns’ (Vihman, 1996, p. 142).

This further development of the concept steps back from any mention of ‘attention’, relying instead on the more embodied idea of an implicit motoric response to a familiar sound pattern, which is hypothesized to create a ‘pop-out’ effect for heard speech forms close to the child’s existing sound patterns. No experimental research had yet been designed to test the concept, however.

3. The Empirical Foundation: Early Production Studies and Vocal Motor Schemes

The basic idea of an articulatory filter grew out of reflections on Ferguson and Farwell (1975), a study that shocked some in the language acquisition community by suggesting that, in the earliest word-learning period, it was not only the meanings of words that drove child attempts at production but also their forms: In producing, relatively accurately, only words with simple adult target forms, children appeared to be selecting words to say on the basis of their sound pattern. But how could an infant know in advance which word forms would prove accessible production targets? The articulatory filter provided a possible answer.

Further circumstantial evidence of an effect of production on perception emerged from a longitudinal production study (seven months of weekly recordings of nine infants: Vihman et al., 1985; see also Oller et al., 1976; Elbers & Ton, 1985) that revealed strong continuity between babble and contemporaneous word forms, with the vocal shapes differing by child. The evidence of such continuity supported the idea that vocal practice might shape early word production, contrary to Jakobson’s long-standing influential account (Jakobson, 1941/1968), which interpreted diary study reports in the light of structuralist theory and held babble (seldom noted in diaries) to be phonetically random and wholly unrelated to word learning.

Vihman and her colleagues’ 1985 study revealed wide individual differences in the early phonetic mastery and timing of consonant production (see also Vihman et al., 1986). Relative use of supraglottal consonants in both babble and words over the period of transition into regular word use proved (for a subset of seven children sampled over three weeks at a comparable lexical level) a good indicator of relative phonological advance at age three (Vihman & Greenlee, 1987), suggesting that ‘the proportion of vocalizations with supraglottal consonants used in the period of transition to speech reflects relative articulatory skill as well as degree of sensitivity to language-like segments and syllables’ (p. 51).

Based on monthly samples from longitudinal studies of the vocal patterns of 20 American infants, McCune and Vihman (1987, 2001) found that, for all but one child, transcription data could be used to identify ‘vocal motor schemes’ (VMS), or frequent, stable

sound patterns that emerge within the babbling period.² Taken as an index of phonetic advance, the number of VMS achieved by each child in this study proved the strongest predictor of age of onset of referential word use, over and above the value of the other measures considered (volubility, proportion of vocalizations with a supraglottal consonant).

In 2010, Keren-Portnoy et al. further demonstrated the value of VMS as a measure of phonetic mastery. This study of 12 British children recorded monthly from 11 months on and then tested on nonword repetition at 26 months showed that the earlier a child had arrived at two VMS, the more advanced was their later memory for just-heard nonword forms.

Together, these production studies established the two premises on which the experimental work reported here would build: that stable individual production routines can be reliably identified in the babble period, and that their timing predicts lexical advance. But it remained unclear whether or how emergent phonetic mastery might affect infants' perceptual experience of input speech. The articulatory filter proposal predicted that infant memory for speech forms would vary in relation to their own output patterns. But how might the hypothesis be tested?

4. Testing the Articulatory Filter: Responses to Speech

In the next few years, several studies drew on VMS as an index of phonetic knowledge to test for an effect of individual vocal patterns on infants' responses to input speech. In the first of these studies to be published, DePaolis et al. (2011) used the Auditory Headturn Preference Paradigm to test 18 British English-learning infants on prepared lists of nonwords loaded with labial, coronal or velar stops, designed to pit, for each infant, pseudowords featuring familiar (VMS) consonants against pseudowords featuring unfamiliar (non-VMS) consonants.

The eighteen infants were recorded repeatedly in their homes until the rapidly transcribed recordings showed that the child had achieved frequent, stable production of at least one VMS. The original stipulation that 10 uses be observed over three sessions was supplemented with a high-frequency criterion (50 uses in one to three sessions), 'to permit timely perceptual testing' (p. 592); nevertheless, it was 'not always... possible to identify the emergence of a single VMS before a second VMS met criterion' (p. 592). Once a VMS was identified, the families were asked to bring the infant in for the experiment.

Unexpectedly, the results proved to be evenly divided between two effects. Half the infants tended to look longer toward the nonword list featuring their VMS consonant, showing a tendency toward the 'familiarity preference' predicted by the articulatory filter proposal. But the remaining nine infants, who had already achieved two VMS, looked significantly longer in response to the *unfamiliar* (non-VMS) pseudoword forms—a 'novelty preference'. Although not the anticipated outcome, both effects supported the general principle that an infant's vocal production practice affects the patterns an infant will find attractive or compelling.

In a follow-up study with Italian infants, Majorano et al. (2014) replicated that split result with a large enough sample (30 infants) to reach significance for both the infants with a single VMS (who showed the familiarity effect) and those with more than one (who showed the novelty effect). Majorano et al. also tested a separate group of (pre-canonical) 6-month-olds, who showed no preference for either set of stimuli, demonstrating that before achieving a first VMS, the infants' vocal forms had yet to be linked with their processing of speech stimuli.

A question that often arises with respect to infant differences in the phonetics of their early vocalizations is the extent to which the speech they are exposed to may shape or constrain them. Vihman et al. (1994) addressed the issue by comparing consonantal occurrence in maternal input with occurrence in the early word forms of five infants each

learning three languages (English, French, Swedish). In all three language groups, the children showed greater variability in their consonant use than did the mothers; there was no evidence of a direct relationship between individual children and their mothers.

Similarly, DePaolis et al. (2011) analyzed consonant use in the child-directed speech of three mothers, both in the session in which two VMS were identified for the infant and in the preceding week. Consistent with Vihman et al. (1994), they found that whereas VMS consonant frequency in all three mothers' IDS was similar, that of the infants was quite diverse and demonstrably unrelated to that of their mothers. Since the mothers proved to be alike while their infants differed, the input cannot account for the differences between the children at this stage of vocal development. (Majorano et al., 2014, undertook similar analyses with similar results.)

DePaolis et al. (2013) carried out a related but differently designed study to test the same concept. Fifty-three British infants (26 learning English, 27 learning Welsh) were recorded in their homes in four bimonthly sessions, from 10.5 to 12 months of age; the number of infant vocalizations that included one of the three stop categories (p/b, t/d, k/g), a nasal (/m, n/) or a fricative (/s/) were tallied for each infant over all four sessions.³ Whereas the earlier studies had tailored the tested contrast to each infant's repertoire, here the contrast was held constant across all infants within each language group, with individual production entered as a continuous measure. The two designs are complementary, the one testing a categorical familiar-unfamiliar prediction, the other a graded relation between production and attention.

Two weeks after the last recording each child was brought in for a perception experiment designed to contrast pseudowords built around two consonants selected for their equivalence in terms of frequency of occurrence in the language in question—for English, /t/d/ vs. /s/, for Welsh, /p/b/ vs. /k/g/—but with individual differences expected for infant production. The Welsh infant production data showed little difference in the frequency of labial vs. velar stops, however, and proved to be unrelated to any infant perceptual 'preference'. For English, in contrast, a significant novelty effect was found, such that the more an infant produced /t/d/ (by far the most commonly occurring consonant category across all sessions), the more they attended instead to /s/-nonwords in the test. Notably, the infants as a group showed no preference for either list, despite the substantial acoustic difference between stops and fricatives; differential attention emerged only as a function of each infant's own production. With the stimuli identical for all infants and matched for frequency in the input, the graded, bidirectional relation between production and attention could not be attributed to any property of the stimuli themselves.

Subsequent to these first tests of the articulatory filter idea, a follow-up study was designed to probe the seeming paradox of familiarity vs. novelty responses in the same experiment (DePaolis et al., 2016). Fifty-nine British English-learning infants were followed longitudinally from 9 months to the point of mastery of two VMS and monthly thereafter to 18 months; at 10 months 53 of them successfully completed the experiment in lexical representation, which contrasted a set of words likely to be familiar by 11 months of age and a phonologically similar set of words unlikely to be familiar ('rare' words: see Hallé & de Boysson-Bardies, 1996; Vihman et al., 2004).

This experiment upended the methodological approach of the earlier studies, as all the infants were brought in to be tested at the same age, regardless of production status (i.e., with or without having achieved two VMS), and the stimuli were real words, not pseudowords. The test age was set to fall a month ahead of the point when most infants have been found to show word-form recognition (Vihman et al., 2007), in the expectation of highly variable responses. The prediction was that (like the six-month-olds responding to pseudowords in Majorano et al., 2014), infants who had not yet begun stable consonant

production (non-VMS infants) would fail to show a preference, while those with one or more VMS would prove a mixed group, showing either a familiarity preference (greater interest in the Familiar words) or a novelty preference (greater interest in the Rare words). This outcome would be in line with the Hunter and Ames (1988) model of infant attention to newly experienced stimuli, which predicts that an initial ‘neutral’ period of no preference for either of two sets of contrasting stimuli will be followed, developmentally, by greater attention, or a stronger ‘preferential’ response, to what has become familiar, and finally, after more exposure or development, a growth of interest in the unfamiliar stimuli at the expense of the familiar.

In other words, the model conceptualizes the direction of an infant’s apparent ‘preference’ as reflecting depth of encoding: none before differential encoding begins, familiarity while it is underway, novelty once it is consolidated. Following this model, the articulatory filter predicts no preference in infants who have not yet achieved stable production patterns, a familiarity response from those with minimal production experience, and a novelty response from those with more consolidated experience. A robust preference in the first group would count against the account. Two caveats are in order, however. Inferences from infant looking times to underlying cognitive processes are far from straightforward (Paulus, 2022), and the Hunter–Ames model itself, though widely invoked, is only now under systematic multi-laboratory evaluation (Kosie et al., 2024). In DePaolis et al. (2016) differential looking is treated as an index of differential processing, with the developmental sequence of preference directions taken as a working interpretive framework rather than established fact.

As anticipated, no overall group difference emerged between the length of looking time elicited by Familiar vs. Rare words. A preferential ratio (*p*-ratio) was calculated to express the proportion of looking toward the Familiar over the total looks to both Familiar and Rare stimuli. The (normally distributed) *p*-ratio continuum was then divided in two: (i) The 26 infants with extreme scores were taken to have ‘passed’ the test by showing some degree of word form recognition, through greater interest in either the Familiar (high *p*-ratio) or the Rare words (low *p*-ratio); (ii) the 27 infants with *p*-ratios closer to the mean of 0.5 (showing little or no evidence of interest in either set of stimuli) were taken to have failed to recognize the familiar words.

The nine-month CDI reports available for 36 of the 59 infants showed a mean of 16 words reported for the Extreme group as compared with 9 for the No-Preference group. Although for this smaller sample the difference did not reach significance, the finding suggested that, as expected, the infants who showed greater interest in either set of words tended to have made a more substantial start on word comprehension than those who failed to show such interest. Finally, growth curves plotted across the data from CDIs for the full nine months of the study showed significantly more reported lexical advance in the Novelty group in comparison with both the No-preference and the Familiarity groups, again in accord with the predictions of the Hunter and Ames model.

Two further studies have since built on VMS as a measure of emergent phonetic mastery to test different effects on word learning. Using a Preferential Looking task, Majorano et al. (2019) demonstrated, with 30 Italian infants aged 11 months, that only those who had achieved at least one VMS could be trained to learn a novel object–novel word form pairing, and only when the pseudowords featured early-learned consonants; non-VMS infants failed at the task, as did VMS-infants when the pseudowords featured late-learned consonants.

Extending the idea of infants’ ‘filtering’ input speech based on their phonetic knowledge or skills, Laing and Bergelson (2020) proposed that, ‘if infants attend preferentially to words that contain their well-established, stable consonants. . . in an experimental set-

ting', they could also be expected to do so 'in an ecologically more realistic setting' (p. 3). That would show that infants not only implicitly pick out from the speech stream word forms made salient by their correspondence with internal representations of their own vocalizations, but also 'begin pairing their VMS consonants with objects they know in their environment, initiating the link between what an infant knows about the world (i.e., object names) and the sounds they can most easily produce. . .' (p. 3).

A database of audio and video recordings of 44 American infants at 10 and 11 months was available to test the idea. From transcription of the most voluble half-hour within each day-long audio recording at 10 months, using the revised criterion of 50 uses of a consonant (DePaolis et al., 2011), 27 infants could clearly be categorized as either With-VMS (N = 13) or No-VMS (N = 14). For the remaining 17, Laing and Bergelson additionally analyzed the transcripts of a half-hour taken from the 11-month all-day audio recording, yielding 11 more With-VMS infants, based on one or both recordings. (These classifications were confirmed in subsequent analysis of vocalizations produced in the video recordings at 10 or 11 months.)

The investigators then tested their hypotheses as to how phonetic knowledge would be reflected in the infants' behavior in their homes. They predicted, first, that With-VMS infants would more often produce babble that matches a caregiver prompt and that they would, in particular, produce their VMS when the caregiver provided such a prompt. Secondly, in a novel extension of the concept, they predicted that when attending to an object, With-VMS infants would respond with a consonant related to the associated name, whereas No-VMS infants would not.

The hypotheses were largely confirmed. Laing and Bergelson found that, regardless of VMS status, infants were more likely than not to include a congruent consonant in vocalizations produced in response to a caregiver prompt (contrary to the less linguistically informed findings of Goldstein & Schwade, 2008). Furthermore, With-VMS infants were more likely to respond contingently to caregiver uses of their within-repertoire consonants than to other caregiver prompts. Finally, at least half the time, With-VMS infants responded with babble that matched the name of an object that held their attention, while No-VMS infants did not. This confirmation of Laing and Bergelson's novel idea suggests that some time before the onset of identifiable word use, phonetically prepared infants may experience an aural 'echo' emanating from often-named items in their everyday environment.

These naturalistic findings bear on the broader question of caregiver feedback. In experimental settings, contingent caregiver feedback has been found to make babble more speech-like (Goldstein & Schwade, 2008). Based on their analysis of recordings made in everyday settings, however, Fagan and Doveikis (2017) reported that caregivers responded to only a minority of infant vocalizations, and infants showed little sign of adjusting their vocal forms to those responses. The articulatory filter proposal identifies the infant's contribution to such exchanges: Caregiver speech and the infant's own vocal practice jointly shape early production, with well-practiced patterns rendering matching input more salient and more likely to be retained.

Lorenzini and Nazzi (2022) pursued the idea that infants' relative vocal advance would affect their speech processing by examining infants' listening preferences for already-familiar word-forms drawn from everyday vocabulary. At both 11 and 14 months, more advanced babblers oriented longer to lists of familiar words containing late-learned consonants than to those with early-learned consonants, showing the novelty preference pattern observed in the earlier pseudoword studies. Under the assumption that already-familiar items are recognized rather than re-encoded at test, the authors concluded that sensorimotor information is retained in the memory traces of stored word-forms and accessed during recognition.

5. Operationalizing Vocal Motor Schemes: Methodological Progress and Open Questions

Over the 60 years since Fry (1966) published his insightful remarks at one of the first psycholinguistics conferences to focus on child language, the value of canonical babble—motoric exploration of syllabic vocalization and the proprioceptive and auditory feedback that goes with it—in laying the articulatory foundation for word learning has been amply demonstrated. (Jakobson’s long-accepted formulation of the orderly, symmetrical emergence of phonological contrasts in early word production, independent of a period of babble, stands as an elegant but misconceived theory-driven view). For the design of experimental procedures to test the role of emergent motor control in the babble and early word period, the construct of VMS, or the identification of stable individual consonantal production preferences from the transcripts of naturalistic infant vocalizations, has proven invaluable.

However, the concept needs further validation. The measure has the advantage of assessing strength (frequency of production of a particular sound pattern) alongside stability (repeated use at a high level over time). But the methodology for specifying emergent VMS remains unmoored and its validity in acoustic terms virtually untested. (For cautionary remarks or critiques of segment-oriented transcription of infant vocalizations, see Holmgren et al., 1986; Rvachew & Brosseau-Lapr , 2018; Rvachew & Alhaidary, 2018.)

The measure has been implemented in a range of different ways, each adapted to the purposes of the particular study and the data at hand. The patterns tested were initially based on 10 infant uses of a given consonant type in three out of four successive half-hour sessions (McCune & Vihman, 1987, 2001; Keren-Portnoy et al., 2010). This was later supplemented or replaced by the selection of consonants showing particularly high frequency use in a single recording (DePaolis et al., 2011; Majorano et al., 2014; Laing & Bergelson, 2020), five uses in two fifteen-minute sessions (Majorano et al., 2019), or even—to assess vocal advance in rare, hard to obtain samples of day-long recordings of children learning Tzeltal (N = 20) or Y l  (N = 12) in rural communities in Mexico and Papua New Guinea, respectively—10 uses over 45 ‘randomly sampled’ minutes (Tzeltal) or 5 uses over 22.5 min (Y l ) (Peute & Casillas, 2022). Although no study has so far addressed the question of the relative validity of these different ways of quantifying VMS, the principle remains the same: To establish a reference point for infant mastery, in the babble period, of one or more supraglottal consonants (or a pair of consonants differing only in VOT) to assess phonetic advance. Validating VMS in acoustic terms matters for the larger argument: An operational index of the articulatory filter should show its stability in the signal itself, independently of transcribers’ judgments.

Accordingly, Cox and Vihman (2024) designed an analysis to ‘observe how acoustic variability develops around the point of VMS attainment’ (pp. 117f.). In this exploratory study testing the validity of transcription-based VMS with direct analysis of the acoustic record, Cox and Vihman analyzed the vocalizations of 28 children recorded in the home once or twice a month from 9 to 18 months (a randomly selected subset of the sample used in DePaolis et al., 2016). VMS had been identified based on the combined criteria of McCune and Vihman (2001) and DePaolis et al. (2011). The point of each child’s achievement of the first of two identifiable VMS was then mapped onto a continuous predictor measured in days to create a timeline along which the acoustic analyses could be laid out. The onset and offset of each stop burst were manually marked, and the primary measure was the spectral centroid at burst release (i.e., the power-weighted average frequency of the burst spectrum, which varies with the size of the cavity in front of the constriction and thus distinguishes stops by place of articulation). Variability in this measure, computed within each place category, could then be tracked over time relative to the point of VMS attainment.

Statistical analysis showed a non-linear decrease in stop production variability over time: An initial period of increasing variability (especially in labials and coronals) was followed by a gradual decrease, with the peak of production variability falling, on average, within 27 days of attainment of two VMS. These findings suggest that a period of exploration is typically followed by stabilization, in keeping with the McCune and Vihman conceptualization of VMS. These are promising results, but further study is needed to provide greater confidence in VMS.

6. Early Canonical and Pre-Canonical Effects of Vocal Experience on Perception of Speech Sounds

The studies described above were primarily intended to evaluate the effect of emergent babble practice on infant responses to and memory for word patterns in the speech stream. A parallel line of research has been concerned with earlier articulatory effects on consonant category formation and discrimination. Recent studies have traced those effects back to the pre-canonical period.

In a study not directly concerned with the articulatory filter idea, [Vilain et al. \(2019\)](#) tested the ability of French infants to categorize place of articulation of heard stops at the cusp of speech-like production. From parental reports, they ascertained whether a child was producing any of 10 consonants (again, the core six stops and two nasals, /s/ or /l/) in multisyllabic babble. The children were then divided into two groups, Babbling (7 6-month-olds, 25 9-month-olds) and Non-babbling (13 6-month-olds only), first, and then two groups based specifically on reported production of the consonants to be tested, /b/ or /d/ (Babbling-b/d, 5 6-month-olds, 23 9-month-olds) and Non-Babbling-b/d (15 6-month-olds, 2 9-month-olds).

The infants participated in an intersensory matching experiment. In baseline trials, they were first presented with side-by-side videos of a woman silently and repeatedly producing the syllable /ba/ in one video and /da/ in the other, and then the videos were shown again, with the sides reversed. A pair of familiarization and test trials followed, exposing the infants first to audio recordings of different female speakers repeatedly producing either /bV/ or /dV/ syllables (where V may be /i, e, u, o/, but not /a/) and then to the /ba/ and /da/ videos. Analysis showed that only Babbling-b/d infants looked longer to the matching video at test than at baseline, indicating that they had generalized the test consonants to new syllabic frames; neither age nor general babbling status proved relevant. The investigators concluded that

The development of production abilities may help infants to refine their perceptual categories and to build reliable phonemic representations. . . When infants start producing a sound, their representation for that sound becomes richer, involving auditory as well as motor information. . . Somatosensory feedback would be integrated with auditory information to form a multisensory representation. ([Vilain et al., 2019](#), pp. 9f.)

This study, then, adds one more type of evidence that infants' response to speech is affected by their emergent vocal practice.

In a series of speech perception studies providing insight into the developmental lead-up to the babbling period, Werker and her colleagues have looked for early production effects on the processing of speech stimuli. [Bruderer et al. \(2015\)](#) were the first to develop a paradigm for detecting an effect of infant articulatory movement on their ability to discriminate unfamiliar sounds, to 'test empirically whether motor processes exert a direct influence on the auditory percept' (p. 13531). These investigators argued that bi-directional effects were plausible: 'Exposure to auditory speech influences the articulatory-motor linkage, and in turn, sensorimotor information from the articulators could impact the

auditory speech percept'. (pp. 13531f.) To rule out possible influence from their preexisting linguistic experiences, Bruderer et al. tested discrimination of a non-English contrast, Hindi dental /ɖ/ versus retroflex /ɖʱ/, which 6-month-olds exposed only to English had been found to discriminate while English-language adults or infants over the age of 10 months could not. Production of these sounds critically requires distinct movements of the tongue tip, forward to the hard palate for /ɖ/, curled back for /ɖʱ/.

To test for an effect of infant articulatory gestures on discrimination of the unfamiliar sounds, the infants were tested with two different teethers, a flat one that effectively prevented tongue-tip motion and a 'gummy tether' that left the tongue free to move. As predicted, infants succeeded in discriminating the two sounds only when lingual movement was unimpeded. The investigators concluded that 'a link between the articulatory-motor and speech perception systems may be more direct than previously thought and is available even before infants accrue experience producing speech sounds themselves' (p. 13535).

Choi et al. (2019) replicated these experiments and achieved the same predicted effect with a native language contrast, /ba/ vs. /da/: A tether designed to prevent lip movement effectively blocked discrimination. In an overview, Choi et al. (2023) argued that although canonical babbling may provide an articulatory filter, 'shaping and refining speech perception' (p. 776),

the sensorimotor mappings for vocal tract articulators available to infants even before they produce their first syllables could interface with the development of the speech and language network and speech experience. (p. 781)

Note, then, that while these studies provide evidence for an auditory-articulatory link before the emergence of canonical babble, they do not speak to the role of actual vocal production or practice in orienting infants' response to speech. For the articulatory filter account, these findings supply the developmental precondition: The sensorimotor linkage on which production-filtered word learning depends is in place before canonical babble provides the child with stable, repeatable targets.

7. Infant Responses to Infant Voices

In a separate but related line of study, some investigators have also looked into infant responses to infant voices. In the first such study, Legerstee et al. (1998) found, experimentally, that 5- and 8-month-old infants listened longer to peer vocalizations than to bells or synthesized notes (a preference for social over non-social stimuli), and to the sounds made by another baby than to their own vocalizations—an apparent novelty effect. Interestingly, however, infants responded *by vocalizing* more in response to their own as compared to another infant's vocalizations. Given the results reviewed here, the vocal response to own vocalizations might be related to the infants' ongoing experience of mapping the sounds they make to the articulatory movements that produce them. In other words, the sound of their own vocalizations may stimulate more vocal exploration; the sounds of other babies, although socially engaging, might not have the same internally motivating effect. In a study designed to deepen our understanding of these findings, Madhavan et al. (2025) are drawing on naturalistic occurrences of vocal play, recorded in the home at ages 4 to 5 months, to further test infant responses to their own and other infants' vocalizations.

In the first of a series of studies, Polka et al. (2014) took up the question of pre-canonical infants' ability to track (synthesized) vowel categories across different talkers, including infants. They found

robust, rapid adjustments [which] may reflect a natural response to the sharp shift in acoustic variability or novelty tied to the infant signals, a perceptual bias favoring infant speech, or some combination of these factors. (p. 1455).

In a follow-up study, extrapolating from the experimental results of DePaolis et al. (2011, 2013) and Majorano et al. (2014), Masapollo et al. (2016) set out to test for an effect of infant production on their listening preferences.

Infants begin producing vowels and vowel-like sounds before they are able to produce the well-formed syllables that characterize canonical babbling . . . If infant ‘output biases intake’ as soon as infants begin to formulate speech sounds, then one would expect to find a perceptual bias favoring infant vowels once vowel production is under way in the pre-babbling stage (at 4–6 months). From this view, one would predict that prebabbling infants would prefer *both* the voice pitch and the formants of infant vowels since these properties jointly form infant speech output. (p. 321)

The experiments confirmed these expectations, providing evidence for what Masapollo et al. term ‘an infant talker bias’ (see also Polka et al., 2022). They see this bias as ‘tied to the development of their vocal self-discovery loop’ (Polka et al., 2025, p. 2)—although it is unclear that a specialized loop need be posited, beyond the automatic feedback vocalizers necessarily receive as they produce sounds. The attraction to infant voices, and the heightened vocal response to one’s own voice, is a further manifestation of the underlying claim reviewed here: that infants’ processing of speech is affected by their proprioceptive and perceptual experience of their own vocal production.

8. Discussion and Conclusions

The primary goal of this overview has been to situate, both developmentally and in terms of the relevant theoretical proposals and tests, the foundation that babbling affords for word ‘selection’ and retention. Studies of the past ten years have provided evidence of the early interaction of articulatory movement and perceptual discrimination, demonstrating the presence of sensorimotor links well before adult-like syllable production. Infant responses to infant vocalizations, perhaps especially their own, provide further evidence of infant preparedness to learn from their own production patterns.

The early motoric advances of the first year lead to the highly consistent emergence, between about 6 and 8 months, of adult-like syllable production. Following Edelman (1987),

we can assume that the child’s own vocal productions leave a ‘sensory-motor neural trace’ which is reactivated and thus strengthened with repeated use. . . Since the trace serves as a receptor for auditory sensation as well as for kinesthetic [and proprioceptive] feedback from motor activity, it will also be strengthened (and perhaps subtly changed or shaped [see Elbers & Ton, 1985]) by adult input patterns which are broadly similar. Thus the child’s vocal patterns flow into a pool of highly familiar auditory patterns, insofar as they ‘match’ adult input. (Vihman, 1993, p. 76)

Pursuing Edelman’s contention that ‘it is the temporal coherence of sensory and motor signals that acts as the selective process’ (Edelman, 1987, pp. 142f.; cf. also Thelen & Smith, 1994, Chapter 5), Vihman (2022) argued that the development of a motor representation (and its associated auditory representation) for an often-repeated vocal form leads to priming by related word forms in input speech, resulting in early word production. In general, those representations should facilitate the mapping and retention of novel word forms and their meanings.

The idea of ‘self-supervised learning’ through an ‘articulatory filter’ has so far been tested based on evidence, from transcribed data, of frequent and stably repeated use of one or more consonants, or VMS. The ‘vocal motor schemes’ reflect a range of individual differences in the identity and timing of the consonants infants produce; this has made it

possible to test, in several experiments, the value of VMS mastery as a measure of infant readiness for lexical growth.

There would surely be uses for such a measure in clinical studies, if its value were to be further confirmed. Existing evidence already points to its predictive value: Earlier VMS attainment predicts lexical advance (McCune & Vihman, 2001; Majorano et al., 2014; McGillion et al., 2017) and stronger nonword repetition at 26 months (Keren-Portnoy et al., 2010). If these findings replicate in larger and more clinically diverse samples, including cochlear-implanted infants, late talkers, and children with developmental language disorders, VMS might prove useful as an early indicator of vocabulary delay. Existing measures in these populations vary widely across studies, limiting cross-study comparison (McDaniel & Gifford, 2020). VMS would offer something more specific: an index of whether a child is developing stable production routines, which would thus capture individual trajectories that stage-based and audiometric measures tend to miss.

It would also be useful to attempt to replicate Majorano et al. (2019), which remains, to our knowledge, the only study to have successfully trained infants to learn novel words, or ‘pseudowords’, as early as 11 months; the critical finding—that only infants who have mastered at least one consonant succeed at the task—should be further tested. Similarly, extension to a larger sample and, ideally, other-language data, of Cox and Vihman’s (2024) painstaking effort to model the emergence of VMS with acoustic analysis would lend greater credibility and substance to the VMS construct. That in turn would strengthen our understanding of the sensorimotor link infants involuntarily construct when they embark on audible jaw-wags, effectively preparing their bid to join the conversation around them.

Author Contributions: Conceptualization, M.M.V.; writing—original draft preparation, M.M.V. and C.M.M.C.; writing—review and editing, M.M.V. and C.M.M.C. All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ See now (Cychosz et al., 2021), which confirms this age range in a large cross-linguistic sample coded for syllable shape by crowdsourcing.
- ² Because infant vowel production is too variable to allow identification of repeated syllabic patterns for the individual child, McCune and Vihman chose to focus on consonants as transcribed in broad IPA. Vowels are more subject to co-articulatory effects and typically receive low reliability scores (Shriberg et al., 1997); finer phonetic distinctions, such as dental vs. alveolar realization, are rarely captured in transcription, as children’s productions often fall between traditional target sounds (Kent & Murray, 1982; Frisch & Wright, 2002); and voicing contrasts have been shown not to be under infant control until months after the period of interest here: see (Macken, 1980). The consonants that qualify as potential VMS are accordingly drawn from a small early-emerging set: stops, nasals, [s] and [l]. Children divide across that set and differ in timing: In McCune and Vihman (2001), all 19 children who achieved VMS by 16 months mastered at least one stop, with nasals, [s] and [l] mastered by only a few each; a similar distribution, split mainly between labial and coronal stops, characterizes the larger samples of DePaolis and colleagues.
- ³ This study was first reported in Vihman and Nakai (2003).

References

- Beckman, M. (1993). Preface: Phonetic development. *Journal of Phonetics*, 2(1/2), 3. [CrossRef] [PubMed]
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influence on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 13531–13536. [CrossRef] [PubMed]

- Choi, D., Bruderer, A. B., & Werker, J. F. (2019). Sensorimotor influences on speech perception in pre-babbling infants: Replication and extension of Bruderer et al. (2015). *Psychonomic Bulletin & Review*, 26, 1388–1399. [CrossRef] [PubMed]
- Choi, D., Yeung, H. H., & Werker, J. F. (2023). Sensorimotor foundations of speech perception in infancy. *Trends in Cognitive Sciences*, 27(8), 773–784. [CrossRef] [PubMed]
- Cox, C., & Vihman, M. M. (2024, June 3–5). *The emergence of stability in infant vocal production: A model based on acoustic analysis*. Fonetik, Annual Swedish Phonetics Meeting, Stockholm, Sweden.
- Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., Scaff, C., Yankowitz, L., & Seidl, A. (2021). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, 24, e13090. [CrossRef] [PubMed]
- DePaolis, R. A., Keren-Portnoy, T., & Vihman, M. M. (2016). Making sense of infant familiarity and novelty responses to words at lexical onset. *Frontiers in Psychology*, 7, 715. [CrossRef] [PubMed]
- 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, 590–601. [CrossRef] [PubMed]
- DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, 36, 642–649. [CrossRef] [PubMed]
- Edelman, G. (1987). *Neural Darwinism: The theory of neuronal group selection*. Basic Books.
- Ejiri, K. (1998). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica*, 55, 226–237. [CrossRef] [PubMed]
- Elbers, L., & Ton, J. (1985). Play pen monologues: The interplay of words and babble in the first words period. *Journal of Child Language*, 12, 551–565. [CrossRef]
- Ertmer, D. J., & Jung, J. (2012). Prelinguistic vocal development in young cochlear implant recipients and typically developing infants: Year 1 of robust hearing experience. *Journal of Deaf Studies and Deaf Education*, 17(1), 116–132. [CrossRef] [PubMed]
- Fagan, M. K. (2015). Why repetition? Repetitive babbling, auditory feedback, and cochlear implantation. *Journal of Experimental Child Psychology*, 137, 125–136. [CrossRef] [PubMed]
- Fagan, M. K., & Doveikis, K. N. (2017). Ordinary interactions challenge proposals that maternal verbal responses shape infant vocal development. *Journal of Speech, Language, and Hearing Research*, 60(10), 2819–2827. [CrossRef] [PubMed]
- Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 51, 419–439. [CrossRef]
- 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. [CrossRef]
- Fry, D. B. (1966). The development of the phonological system in the normal and the deaf child. In F. Smith, & G. A. Miller (Eds.), *The genesis of language: A psycholinguistic approach. Proceedings of a conference on 'language development in children'* (pp. 187–206). MIT Press.
- Goldstein, M. H., & Schwade, J. A. (2008). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Psychological Science*, 19, 515–523. [PubMed]
- Hallé, P., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, 19, 463–481. [CrossRef]
- Holmgren, K., Lindblom, B., Aurelius, G., Jalling, B., & Zetterstrom, R. (1986). On the phonetics of infant vocalization. In B. Lindblom, & R. Zetterström (Eds.), *Precursors of early speech* (pp. 51–63). Macmillan Press.
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier, & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 5, pp. 69–95). Greenwood Publishing Group.
- Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, 37(2), 229–261. [CrossRef] [PubMed]
- Iverson, J. M. (2021). Developmental variability and developmental cascades: Lessons from motor and language development in infancy. *Current Directions in Psychological Science*, 30(3), 228–235. [CrossRef] [PubMed]
- Jakobson, R. (1968). *Kindersprache, Aphasie und Allgemeine Lautgesetze* [Child language, aphasia, and phonological universals]. Mouton. (Original work published 1941)
- 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. [CrossRef] [PubMed]
- Keren-Portnoy, T., Vihman, M. M., DePaolis, R., Whitaker, C., & Williams, N. A. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language, and Hearing Research*, 53, 1280–1293. [CrossRef] [PubMed]
- Koopmans-van Beinum, F. J., & Van der Stelt, J. M. (1986). Early stages in the development of speech movements. In B. Lindblom, & R. Zetterström (Eds.), *Precursors of early speech* (pp. 37–50). Macmillan Press.
- Kosie, J. E., Zettersten, M., Abu-Zhaya, R., Amso, D., Babineau, M., Baumgartner, H., Bazhydai, M., Belia, M., Benavides-Varela, S., Bergmann, C., Berteletti, I., Black, A., Borges, P., Borovsky, A., Byers-Heinlein, A., Cabrera, L., Calignano, G., Cao, A., Chijiwa, A., . . . Lew-Williams, C. (2024). *ManyBabies 5: A large-scale investigation of the proposed shift from familiarity preference to novelty preference in infant looking time*. PsyArXiv. Available online: https://osf.io/preprints/psyarxiv/ck3vd_v1 (accessed on 1 September 2026).

- Laing, C., & Bergelson, E. (2020). From babble to words: Infants' early productions match words and objects in their environment. *Cognitive Psychology*, 122, 101308. [\[CrossRef\]](#) [\[PubMed\]](#)
- Legerstee, M., Anderson, D., & Schaffer, A. (1998). Five- and eight-month-old infants recognize their faces and voices as familiar and social stimuli. *Child Development*, 69(1), 37–50. [\[CrossRef\]](#)
- Locke, J. L. (1986). Speech perception and the emergent lexicon: An ethological approach. In P. Fletcher, & M. Garman (Eds.), *Language acquisition: Studies in first language development* (2nd ed., pp. 240–250). The University Press.
- Lorenzini, I., & Nazzi, T. (2022). Early recognition of familiar word-forms as a function of production skills. *Frontiers in Psychology*, 13, 947245. [\[CrossRef\]](#) [\[PubMed\]](#)
- Macken, M. A. (1980). Aspects of the acquisition of stop systems: A cross-linguistic perspective. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, I: Production*. Academic Press.
- Madhavan, R., Blake, C., Oxley, F., Smith, K., & Laing, C. E. (2025). *Do infants prefer voluntary vocalisations from their own vocal tract over those of other infants?* Registered Report.
- Majorano, M., Bastianello, T., Morelli, M., Lavelli, M., & Vihman, M. M. (2019). Vocal production and novel word learning in the first year. *Journal of Child Language*, 46, 606–616. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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, 179–204. [\[CrossRef\]](#)
- Masapollo, M., Polka, L., & Ménard, L. (2016). When infants talk, infants listen: Pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*, 19, 318–328. [\[CrossRef\]](#) [\[PubMed\]](#)
- McCune, L., & Vihman, M. M. (1987). Vocal motor schemes. *Papers and Reports on Child Language Development*, 26, 72–79.
- McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development. *Journal of Speech, Language and Hearing Research*, 44, 670–684. [\[CrossRef\]](#) [\[PubMed\]](#)
- McDaniel, J., & Gifford, R. H. (2020). Prelinguistic vocal development in children with cochlear implants: A systematic review. *Ear and Hearing*, 41(5), 1064–1076. [\[CrossRef\]](#) [\[PubMed\]](#)
- McGillion, M. M., Matthews, D., Herbert, J., Pine, J., Vihman, M. M., Keren-Portnoy, T., & DePaolis, R. A. (2017). What paves the way to conventional language? The predictive value of babble, pointing and SES. *Child Development*, 88, 156–166. [\[CrossRef\]](#) [\[PubMed\]](#)
- Meier, R. P., McGarvin, L., Zakia, R. A. E., & Willerman, R. (1997). Silent mandibular oscillations in vocal babbling. *Phonetica*, 54(3–4), 153–171. [\[CrossRef\]](#) [\[PubMed\]](#)
- Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, I: Production* (pp. 93–112). Academic Press.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Lawrence Erlbaum.
- Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59, 441–449. [\[CrossRef\]](#)
- Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, 3, 1–11. [\[CrossRef\]](#)
- Paulus, M. (2022). Should infant psychology rely on the violation-of-expectation method? Not anymore. *Infant and Child Development*, 31(1), e2306. [\[CrossRef\]](#)
- 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), 1–28.
- Polka, L., Alonso-Arteche, M. F., Phillips, N. K., Moradi, S., Ménard, L., & Masapollo, M. (2025). Infants' attraction to infant vocalizations. *Infant Behavior and Development*, 81, 102150. [\[CrossRef\]](#) [\[PubMed\]](#)
- Polka, L., Masapollo, M., & Ménard, L. (2014). Who's talking now? Infants' perception of vowels with infant vocal properties. *Psychological Science*, 25, 1448–1456. [\[CrossRef\]](#) [\[PubMed\]](#)
- Polka, L., Masapollo, M., & Ménard, L. (2022). Setting the stage for speech production: Infants prefer listening to speech sounds with infant vocal resonances. *Journal of Speech, Language, and Hearing Research*, 65, 109–120. [\[CrossRef\]](#) [\[PubMed\]](#)
- Roug, L., Landberg, I., & Lundberg, L.-J. (1989). Phonetic development in early infancy: A study of four Swedish children during the first eighteen months of life. *Journal of Child Language*, 16, 19–40. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rvachew, S., & Alhaidary, A. (2018). The phonetics of babbling. In M. Aronoff (Ed.), *Oxford research encyclopedia of linguistics* (online ed.). Oxford Academic.
- Rvachew, S., & Brosseau-Lapr  , F. (2018). *Developmental phonological disorders: Foundations of clinical practice* (2nd ed.). Plural Publishing.
- Rvachew, S., Slawinski, E. B., Williams, M., & Green, C. L. (1999). The impact of early onset otitis media on babbling and early language development. *Journal of the Acoustical Society of America*, 105, 467–475. [\[CrossRef\]](#) [\[PubMed\]](#)
- Schauwers, K., Gillis, S., Daemers, K., De Beukelaer, C., & Govaerts, P. J. (2004). Cochlear implantation between 5 and 20 months of age: The onset of babbling and the audiologic outcome. *Otology & Neurotology*, 25(3), 263–270. [\[CrossRef\]](#) [\[PubMed\]](#)
- Shriberg, L. D., Austin, D., Lewis, B. A., McSweeney, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research*, 40(4), 708–722. [\[PubMed\]](#)
- Stark, R. E. (1980). Stages of speech development in the first year of life. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, I: Production* (pp. 73–92). Academic Press.

- Thelen, E. (1981). Rhythmical behavior in infancy. *Developmental Psychology*, 17, 237–257. [\[CrossRef\]](#)
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press.
- Vihman, M. M. (1991). Ontogeny of phonetic gestures: Speech production. In I. G. Mattingly, & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception*. Lawrence Erlbaum Associates.
- Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21(1/2), 61–82. [\[CrossRef\]](#)
- Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Basil Blackwell.
- Vihman, M. M. (2019). *Phonological templates in development*. Oxford University Press.
- Vihman, M. M. (2022). The developmental origins of phonological memory. *Psychological Review*, 129, 1495–1508. [\[CrossRef\]](#) [\[PubMed\]](#)
- Vihman, M. M., Ferguson, C. A., & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*, 7, 3–40. [\[CrossRef\]](#)
- Vihman, M. M., & Greenlee, M. (1987). Individual differences in phonological development: Ages one and three years. *Journal of Speech and Hearing Research*, 30, 503–521. [\[PubMed\]](#)
- Vihman, M. M., Kay, E., de Boysson-Bardies, B., Durand, C., & Sundberg, U. (1994). External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to one-year-old children. *Developmental Psychology*, 30, 652–663. [\[CrossRef\]](#)
- Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A reassessment of the continuity issue. *Language*, 61, 397–445. [\[CrossRef\]](#) [\[PubMed\]](#)
- Vihman, M. M., & Nakai, S. (2003). Experimental evidence for an effect of vocal experience on infant speech perception. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences: ICPhS, Barcelona* (pp. 1017–1020). ICPhS Organizing Committee.
- Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50, 336–353. [\[CrossRef\]](#)
- Vihman, M. M., Thierry, G., Lum, J., Keren-Portnoy, T., & Martin, P. (2007). Onset of word form recognition in English, Welsh and English-Welsh bilingual infants. *Applied Psycholinguistics*, 28, 475–493. [\[CrossRef\]](#)
- Vilain, A., Dole, M., Loevenbruck, H., Pascalis, O., & Schwartz, J.-L. (2019). The role of production abilities in the perception of consonant category in infants. *Developmental Science*, 22, e12830. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.