The developmental origins of phonological memory

Marilyn M. Vihman
University of York

mv509@york.ac.uk

Dedication

I dedicate this paper to the memory of Bill Macken (Cardiff University), whose death in February 2020 came as a shock when I learned of it, just as I began work on this paper, in September of that year. At the 2012 Student Congress on Cognitive Science in Umeå, where both Bill and I gave talks, he introduced me to the idea, which he and Dylan Jones had developed over their many years of experimental work with adults, that 'all cognition is sensory-motor processing' (p.c., 2012). This paper owes a great deal to the critical new perspective on phonological development that I gained from that fortunate encounter with Bill and the discussions that followed.

Acknowledgement: I thank my colleagues Chris Cox, Catherine Laing, Lorraine McCune and Lise Menn for their helpful comments on an earlier draft.

Abstract

Phonological memory, or the ability to remember a novel word string well enough to repeat it, has long been characterized as a time-limited store. An alternative embodiment model sees it as the product of the dynamic sensorimotor (perceptual and production) processes that inform responses to speech. Keren-Portnoy et al. (2010) demonstrated that this capacity, often tested through nonword repetition and found to predict lexical advance, is itself predicted by the first advances in babbling. Pursuing the idea that phonological memory develops through vocal production, I trace its development – drawing on illustrative data from children learning six languages – from the earliest adult-like vocalizations through to the first words and the consolidation of early words into an initial lexical network and more stable representational capacity. I suggest that it is the interaction of perceptual and production experience that mediates the mapping of new forms onto lexical representations.

The goal of this review is to clarify the developmental foundations of phonological memory and its role in word learning by tracing its emergence back to infants' first adult-like vocalizations and early word production. I will claim that phonological memory develops, to a large extent, along with and as a result of practical experience with vocal production. Furthermore, I take differences in the maturation of articulatory control and the extent of infants' vocal practice and phonetic skill, along with likely differences in their (perceptual) sensitivity to matches between heard speech and patterns familiar from their own vocal output, to be at the root of differences in non-word repetition or NWR, a measure of phonological memory sometimes confused with the construct itself. In short, I will argue that *vocal production builds phonological memory,* providing a basis for mapping a new form onto an initial representation (a first step in word learning, or 'lexical configuration': Leach & Samuel, 2007), while repeated word use permits consolidation ('lexical engagement': Leach & Samuel, 2007). This second step leads to the emergence of a lexical network and more stable representational capacity. Lexical growth thus results in increasing phonological knowledge and, accordingly, in more robust, better articulated phonological memory and more efficient word learning. In the final section I briefly review two competing theoretical models of working memory to ask whether a cognitive model such as that of Baddeley and Gathercole can account for phonological memory in its earliest manifestations or whether the very different embodiment model proposed by Jones and Macken (e.g., Macken & Jones, 2003; Jones, Macken & Nicholls, 2004; D. M. Jones & Macken, 2018), with its emphasis on direct sensorimotor interaction, offers a more plausible conceptualization that can also readily be related to a developmental perspective.

Phonological memory, or the ability to remember a novel word string well enough to repeat it, has been the subject of considerable research for over thirty years, most of it couched within Baddeley's (1986) working memory framework. Gathercole and Baddeley (1989) were the first to attempt to establish that what Baddeley termed the 'articulatory loop component of working memory', later referred to as the 'phonological loop', serves to support vocabulary learning in young children. Baddeley, Gathercole and Papagno (1998) provide a useful review of the considerable experimental literature on the phonological loop available by that time, based on studies of adults, children, and neuropsychological patients. These studies identified as the function of the phonological loop the provision of support for learning novel word forms.

Gathercole, in particular, has championed non-word repetition (NWR) as a key test of children's phonological memory[1] and has demonstrated the value of these tests in predicting lexical

[1] Digit span is an alternative measure of verbal short-term memory, with the important difference that digits are highly familiar lexical units, so that recall of a list of digits necessarily draws on long-term memory while NWR is intended as a 'purer' measure of short-term memory (Gathercole & Baddeley, 1989). G. Jones and Macken (2018) provide comparative analyses of the outcomes of tests involving digit lists, nonword lists and mixed digit and word lists, based on both computational modelling and behavioral testing with 6-year-olds. Because the use of NWR has been so prominent in studies of child language development, I restrict myself to that literature here, but G. Jones and Macken forcefully argue that long-term associative learning is sufficient to account for the effects seen in either digit span or NWR as measures of verbal short-term memory.

advance in numerous papers (see Gathercole, 2006). A central assumption of the approach is that what constrains children's ability to repeat nonwords is the necessity to hold a phonological representation in (time-limited) working memory in order to be able to articulate it:

> Auditory linguistic inputs are automatically represented in the store, where they are subject to rapid time-based decay. The decay of the representations can be offset by a subvocal rehearsal process that boosts their activation levels. Rehearsal is a volitional strategy that is closely associated with covert articulatory processes and that does not typically emerge until after 7 years of age. (Gathercole, 2006, p. 519).

Thus the representation that supports new word learning is conceived of as taking up space in a time-limited store that can be refreshed through subvocal rehearsal – but such rehearsal would obtain only in school-age children. This implies that, for younger children, differences in NWR scores must reflect differences in phonological storage capacity. Specifically, Gathercole (2006) proposes that the phonological store be conceptualized as 'the set of currently activated phonological representations' and goes on to clarify what she sees as the source of individual differences in performance on NWR:

> The quality of the representations at the point of retrieval is influenced both by factors operating at perceptual analysis that determine the quality of the phonological representations (e.g., acoustic quality and phonotactic frequency), and by the endurance of these representations over time. Variation in phonological storage capacity between individuals may result from either differences in initial encoding or endurance, or both. The precise nature of variation in endurance is unknown at present, but may relate to differences in the rate of decay of phonological representations and/or of their resistance to interference. (p. 522)

I should add that both Baddeley et al. (1998) and Gathercole (2006) make clear that they also see prior lexical knowledge as supporting phonological memory in children to some extent:

> Individual differences in children's S[hort T[erm] M[emory] performance prove to be related to their vocabulary knowledge, with children who perform well on tests of verbal STM typically also having good vocabulary knowledge…It is in principle as plausible that good vocabulary knowledge supports accurate nonword repetition as the reverse. (Baddeley et al., 1998, p. 159f.)

In fact, Baddeley et al. (1998) see the role of prior knowledge as offsetting the 'capacity constraints' of the phonological loop. They suggest that '*imitating the sounds of new words may be a natural strategy* that serves to boost vocabulary acquisition by enhancing phonological loop representations of the novel phonological structures' (p. 162, italics added), but they do not consider the importance of production more generally as a potential source of support for retaining novel forms. Baddeley et al. argue that, early on, 'differences in the capacity to repeat back unfamiliar items will lead to differences in vocabulary, which *only subsequently begin to have a reciprocal influence on nonword repetition performance'* (p. 169, italics added). Interestingly, however, despite the suggestion that a system that functions to support word learning should be able to make use of prior phonological knowledge, these authors also warn that it should not be restricted by that knowledge, lest it prove 'insensitive to novel inputs':

> What is required therefore is some form of *temporary representation* that can both provide an accurate if brief record of specific potentially novel input while relating that input to the long-term system that represents the prior knowledge of the structure of language. The system thus needs some form of *temporary activation'* (p. 169, italics added).

More succinctly, and in more direct contrast to what I will argue here, Gathercole (2006) interprets 'the developmental association between nonword repetition and vocabulary…as a consequence of the common involvement of phonological storage in both activities' [i.e., 'both the representation and learning of new words]' (p. 529).

It will be the burden of this paper to explore the sources of activation of prior knowledge in new word learning; contrary to Baddeley et al. (1998), the effect of previous phonological experience will be taken to be long-lasting and on-going rather than temporary.

Challenges to the model

The validity of NWR as a measure of working memory capacity was challenged early on. Following up on one of the methodological issues raised in Snowling, Chiat and Hulme's (1991) commentary on Gathercole, Willis, Emslie and Baddeley (1991), Dollaghan, Biber and Campbell (1995) pointed out that

> very little is known about the processes that support performance on nonsense word repetition tasks. At a minimum, the successful performance of such tasks seems to require that the listener encode acoustic-phonetic strings as phonemes and syllables, mark these units according to their temporal sequence, and use this representation of phoneme sequences to guide the preparation and execution of a motoric program for articulating the string. (pp. 211f.).

However, nonwords have been found to prime phonologically similar lexical items (e.g., Connine, Blasko & Titone, 1993). To systematically test for a lexical effect on NWR performance in children, Dollaghan et al. presented 30 schoolboys aged 9 to 13 with 3- or 4-syllable nonwords that did or did not include a monosyllabic English word in the stressed-syllable position. The study found clear lexicality effects, both positive (better overall accuracy in repeating long words that include a short real word) and negative (replacing unfamiliar stressed syllables with corresponding real words), thereby casting doubt on the validity of NWR as a 'pure' measure of short-term memory.

Pursuing the question of the role of existing knowledge on NWR performance, Metsala (1999) specifically contested the idea that NWR plays a causal role in vocabulary growth, maintaining instead that as the number of known words increases, 'lexical restructuring' – more specifically, segmentalization, or the specification of segmental sequences in lieu of more holistic representations – ensues, driven by the need to maintain contrast (see also Fowler, 1991; Walley, 1993); in this view, it is segmentalization that facilitates NWR performance.

In a longitudinal study of 71 children seen at ages 5 and 6 years Bowey (2001) weighed evidence for the working memory capacity model as compared with lexical restructuring. Supporting the latter, Bowey concluded that

> the concurrent association between nonword repetition and vocabulary is mediated by an underlying phonological processing ability, which may partly reflect the quality of phonological representation available within the language module. (p. 463)

The effect of prior lexical knowledge has continued to be tested experimentally over the years, with conflicting outcomes. In their study of 54 children with Specific Language Impairment aged between 5 and 10 years and matched controls Edwards & Lahey (1998) found that the test correlated significantly with expressive but not with receptive language measures. In contrast, in a study of 86 typically developing four-year-old children Cychosz, Erskine, Munson & Edwards (2021) found, based on accuracy of repetition of just the initial (stressed) CV sequence, a marginally better fit with receptive vocabulary. However, in a longitudinal Norwegian study of 219 children aged 4 to 7 years – a replication and reanalysis of Gathercole, Willis, Emslie & Baddeley (1992) – Melby-Lervåg, Lervåg, Lyster, Klem, Hagtvet & Hulme (2012) failed to find a significant effect of either NWR on vocabulary size or the reverse, leading these authors to question whether the phonological loop actually serves to support language learning at all.

The ‘quality of phonological representations’ comes up as a way to account for differences in children’s performance on the NWR task in several of these studies. In reality, as Munson (2006) points out, the nature or quality of phonological representations can only be guessed at or indirectly inferred. Baddeley et al. (1998) comment that

> a system in which learning occurs incrementally over time, on the basis of the detection of repeated features of temporary memory representations, would allow a long-term record of new words to be based on abstractions of sound patterns consistent over several exposures (p. 169).

Since then, a good deal of research (e.g., Hawkins, 2003; Johnson, 2007; Pierrehumbert, 2016, among others) has provided evidence to support the somewhat differently conceived but compatible view that representations take the form of ‘detailed memories of individual perceptual “episodes”’ (Munson, 2006, p. 579) – in other words, of exemplars, whose accumulation over time gradually permits segmental representations to be abstracted out of their lexical contexts (Munson, Edwards & Beckman, 2012).

The exemplar model fits well with a holistic interpretation of early phonological representations, which, in the everyday life of a toddler, are necessarily dynamic, reflecting the cumulative (but variable) effect of multiple encounters with a word, produced by different speakers and embedded in different situational contexts (cf. Menn, Schmidt & Nicholas, 2013; for a model of holistic representations in adult speech production, see Davis & Redford, 2019). The concept of holistic representation suggests that, in the early word-learning period, the child may retain the overall shape and salient elements of a word form rather than its complete segmental sequence (see Vihman, Ota, Keren-Portnoy & Lou, in revision). Relative salience will depend on many factors, including prosodic structure (Snowling et al., 1991) and, critically, the child’s previous experience with (a) the form as a whole and its subcomponents in input speech (i.e., the effect of both the frequency of occurrence of the word and the phonotactic probabilities of its subcomponents: cf. Cychosz et al., 2021; G. Jones, Cabiddu, Andrews & Rowland, 2021) and also (b) the child’s own production practice (Vihman, 1993; Keren-Portnoy, Vihman, DePaolis, Whitaker & Williams, 2010; Menn et al., 2013).

Cychosz et al. (2021) argue that if, as the holistic view would maintain,

> children rely on chunks to acquire speech categories, children should find nonwords more difficult to process and repeat because children do not have any memory traces of the nonwords. As a result, they cannot rely on entire episodic traces during lexical retrieval and production. Furthermore, this theory predicts that even though subcomponents of nonwords may have some mental representation – because they form parts of real words that the children know – those representations are likely fainter and the memory traces less established because the sequences are less frequently heard and produced. (p. 35)

In the event, Cychosz et al. found that the four-year-olds in their study ‘repeat CV sequences more accurately when the sequences are contained in real words rather than nonwords’ (p. 47), but also noted that

> the fact that the children successfully repeated the nonwords at all is actually evidence that the participants do have a segmental, or at least sub-lexical, phonology. If there were no abstraction at all from the original lexical context, then children would be entirely incapable of repeating nonwords’ (p. 49).

Indeed, more abstract or segmentalized representations of speech must be assumed to emerge over time under any interpretation. However, Noiray, Popescu, Killmer, Rubertus, Krüger and Hintermeier (2019), based on an impressive longitudinal ultrasound study of co-articulation in 41 children aged 4 to 7, provide evidence that casts some doubt on the lexical restructuring hypothesis. These authors find that ‘from very early in development, the process of coarticulation itself binds gestures, sounds, phonetic units together to create compounds that ultimately lend meaning to speech streams’ (p. 12); rather than segmentalization being the child’s response to pressure for contrast in an increasingly dense lexicon, ‘holistic and segmental organizations (both in representation and production) develop together, albeit probably at different paces at different times’ (p. 12). That is, early holistic representations are gradually superseded, in a dynamic, non-linear process, by changes to the specification of sub-lexical units, while

> motor, lexical, and phonological developments collaborate dynamically over time by contact with the language (i.e., via increasingly richer exposure and *practice speaking the language*)’ (Noiray et al., 2019, p. 13, italics added).

From this perspective, performance on verbal short-term memory tests can be taken to reflect the on-line activation of transient representations that draw heavily**,** especially in the early years of word learning, on the child’s experiences of hearing the word, or subcomponents of it, and of producing all or parts of it – that is, on the perceptual and production experiences encoded in their emergent lexicon. In accord with this view, I will argue here that ‘phonological memory’ is best understood to refer not to a ‘store’ of any kind but to the highly dynamic processes that inform individual responses to speech (cf. Macken & Jones, 2003).

Studies of infants and toddlers

The early studies of phonological memory tested children aged 3 at the youngest and tended not to consider the possible sources of individual differences in NWR – that is, whether they reflect an inherent characteristic of the child, as is sometimes implied, or are rather the product of differences that arise over the course of language development. However, Parra, Hoff and Core (2011) showed that, contrary to the notion of inherent or ‘biological’ differences, a bilingual child's phonological memory, as tested separately with NWR in each language, may depend on the relative strength of the child's lexical knowledge *in the language being tested* (see also Conboy & Mills, 2006; Marchman et al., 2010; for a review of bilingual studies and the variable relation of phonological memory to children's relative lexical and grammatical advance in their two languages, see Kehoe, Poulin-Dubois & Friend, in press). This is in accord with evidence that phonological memory – especially in preschoolers – not only may predict lexical advance but also reflects vocabulary size (Metsala, 1999; Verhagen, Boom, Mulder, de Bree & Leseman, 2019; Cychosz et al., 2021). Thus it is worth taking seriously the possibility that phonological memory emerges from the learning processes involved in lexical development.

More recent research has addressed considerably younger children than were tested in the early studies. Roy and Chiat (2004) developed and tested a prosodically sensitive NWR task with two- as well as four-year-olds; what must be the largest study carried out to date (a longitudinal study of 471 children) covered ages two to five years (Verhagen et al., 2019). Hoff, Core and Bridges (2008) showed that NWR can provide reliable results even with children under age two and found that relative success in the task is related to the child's expressive vocabulary size at the time of testing.

Keren-Portnoy et al. (2010) demonstrated a strong association between individual differences in age at first consonant mastery in the prelinguistic period and phonological memory as assessed with NWR at age two: The earlier a child showed strong and stable use of two or more supraglottal consonants (between 11 and 17 mos., mean age 12 mos.), the higher their score on NWR at 26 months. In other words, the individual differences in early phonetic skills that can be observed in naturalistic studies of babble and first words later translate into differences in 'phonological memory skills' as tested through NWR. We can infer that the early word learning and use that reliably follows the first stabilization of consonant production serves as a key mediator between those early phonetic skills and later phonological memory.

The Keren-Portnoy et al. study, though limited to only 12 children followed longitudinally, provided a novel insight into the nature of phonological memory.

> The correlation between the age at controlled production of consonants and performance on the nonword repetition test suggests that phonological working memory is built through experience with use of well-categorized segments or syllables: The longer a child has been producing well-controlled consonants, the better developed is his or her phonological working memory. (p. 1288).
>
> It is possible that the causality actually operates in the opposite direction or is bidirectional: Perhaps children who have good sensitivity to and memory for sound patterns will be better at learning to modulate their babble...Our correlational data do not allow us to distinguish between these interpretations. What is clear, however, is that sensitivity to and memory for sound patterns and advances in accurate consonant production develop hand in hand and are interdependent (p. 1289).

I follow up on these ideas here, tracing the emergence of stable consonant use in the first year of life and of systematic phonological representations thereafter. I will return to question the interpretation of phonological memory as a 'store' in the concluding section.

Word learning and system building

Emergence of speech-like vocalizations and effects of the ambient language

Adult-like 'canonical' or CV syllables emerge on a maturational schedule, between about 6 and 8 months (with an inability to hear input speech the factor most likely to result in delay or failure to produce babble at all: Oller, 2000). This is consistent with studies that point to the syllable as a basic foundation for the uniquely human ability to produce speech (MacNeilage, 1998) as well as for its perceptual processing, which humans share with many phylogenetic predecessors (Hauser, 1996), and, crucially, for their dynamic interaction (Poeppel & Assaneo, 2020). Word use is observable only a few months after the onset of canonical babbling (Fagan, 2009) – likely because mastery of one or more supraglottal consonants is a prerequisite skill, an advance that has been defined in terms of Vocal Motor Scheme use (VMS: at least one consonant repeatedly and stably produced, as shown either in high incidence within a single half-hour recording or regular use in two or more recordings spaced days or weeks apart [McCune & Vihman, 2001; Keren-Portnoy et al., 2010; Laing & Bergelson, 2020]; differences in voicing are disregarded for this measure, as infants begin to control voicing only several months later: Macken, 1980).

While few if any effects of the ambient language have been reported for the first canonical syllable production, ongoing exposure to adult speech leads to unconscious shaping of infant vocalizations toward frequently occurring phonetic elements of the adult language that also fall within the infants' production capacity. That is, babbling and early words are typically no more than one or two syllables in length – but monosyllables are more common in English than in French or Swedish, for example (Vihman, Kay, Boysson-Bardies, Durand & Sundberg, 1994). Similarly, the stops, nasals, glottals and glides that constitute the onsets to the first canonical syllables are affected by the frequency of occurrence of those consonant types in the speech infants hear (with more labials in some languages, coronals in others, for example: Boysson-Bardies & Vihman, 1991). The most common vowels are low and front or central, but with a tendency, within that restricted range, for the preferred vowel space of the ambient language to be best represented in babbling (Boysson-Bardies, Hallé, Sagart & Durand, 1989; see also Rvachew, Alhaidary, Mattock & Polka, 2008). And children acquiring Mandarin produce in their babbling not only a falling tone, a contour widely reported for babbling in English as well (Kent & Murray, 1982), but also the motorically accessible and commonly occurring high-level tone, which children exposed only to English do not tend to produce (Lou, 2021). Complementarily, the phonetic types and prosodic structures that are typically learned only later (e.g., word forms of more than two syllables, codas, fricatives, affricates or liquids, high and back vowels) generally begin to be observed only some weeks after word use has become well established, and then only in languages in which those types or structures occur with high frequency. This reflects a shift in production, as vocal forms come to be directed at specifically targeted adult words or short phrases and the child's articulatory skills gradually expand to accommodate a growing range of diverse targets (Vihman & Boysson-Bardies, 1994).

Thus, while the concept of 'babbling drift' (Brown, 1958), or a gradually increasing prelinguistic influence of the sound structure of the input on the child's vocal forms, has continued to be disputed by some (see, for example, Lee, Jhang, Chen, Relyea & Oller, 2017), an impact of the phonetic characteristics of the ambient language on the child's vocal production in the period of transition from babble to words is well documented. However, it is important to keep in mind that the child's production capacity limits the extent of any such impact. Indeed, the evidence of a subtle influence of perceptual experience on vocal production in this early period is best interpreted as suggesting that infants respond to phonetic elements in input speech that roughly correspond to the vocal forms they themselves are producing by preferentially repeating or reusing those elements of their own repertoire (see Westermann & Miranda, 2004).

Effects of vocal production on infant processing of input speech and first word learning

In this same period lexical comprehension is slowly emerging (Bergelson & Swingley, 2012), suggesting a growing ability of the child to retain not only sound patterns (Jusczyk, 1997) and word forms (Hallé & Boysson-Bardies, 1994) but also their situational context, although word production may not yet have begun. Word form recognition is a prerequisite for comprehension, as situational meanings can only be robustly related to word forms if the latter are recognizable on repeated encounters in the input (Swingley, 2009). It has now been well established experimentally that by 11 months most infants can recognize word forms to which they are frequently exposed in everyday life, without training in the lab (in French: Hallé & Boysson-Bardies, 1994, English: Vihman, Nakai, DePaolis & Hallé, 2004, Dutch: Swingley, 2005, Italian: Vihman & Majorano, 2017, Hebrew: Segal, Keren-Portnoy & Vihman, 2020); at 9 months a group effect is not yet apparent behaviourally (Vihman et al., 2004), although some evidence of a brain response can already be detected at that age (Thierry & Vihman, 2008).

The onset of adult-like syllable production a few months before the age of first untrained word-form recognition provides circumstantial evidence that vocal practice helps to focus infant attention on word forms in the input and thus to facilitate retention of at least a few commonly used word forms (Vihman, 2002). Experimental studies have confirmed this hypothesis: Once infants have met the operational criteria for VMS they show differential interest in the consonant(s) they themselves are or are not yet using: Infants observed to be using more than one VMS show greater interest in nonword stimuli built around a consonant *not* yet in repertoire (DePaolis, Vihman & Keren-Portnoy, 2011, DePaolis, Vihman & Nakai, 2013, Majorano, Vihman & DePaolis, 2014), while infants with only a single VMS attend most to stimuli that include a consonant familiar from their own use (Majorano et al., 2014). (For example, the Italian stimuli in Majorano et al. included /pabʌ/ and /kɔgu/, each of which could be expected to be familiar to some infants but not to others.) Furthermore, a study with 59 infants aged 10 months – that is, a month before infants are generally found to 'succeed' on the word-form recognition task (i.e., to distinguish familiar from unfamiliar word forms) – further supported the idea: Infants who had shown strong and stable use of two consonants before being tested attended longer to either the familiar words (a familiarity effect) or the unfamiliar words (a novelty effect), while infants who had not attained the two VMS criterion failed to show differential attention to either set of stimuli (DePaolis, Keren-Portnoy & Vihman, 2016).

In a complementary finding that supports the link with emergent language comprehension Laing and Bergelson (2020) report that, in their sample of 44 infants recorded in naturalistic play with a caregiver, those who had met VMS criteria were significantly more likely to produce a vocal form that matches a consonant in the name of an object to which they were attending than infants who had not met the criteria. This suggests that only infants with sufficient (expressive) vocal resources had begun to form phonological representations for the labels of common objects. In more direct relation to word learning, McGillion, Matthews, Herbert, Pine, Vihman, Keren-Portnoy & DePaolis (2017) found, based on naturalistic recordings of 48 infants at play with their caregiver, that onset of stable consonant production (or VMS) was a better predictor of first word production than either the family's socio-economic status or the onset of pointing. And, finally, in an experimental study Majorano, Bastianello, Morelli, Lavelli & Vihman (2019) recorded 30 Italian infants to establish whether they had attained one or more VMS; the investigators then had the children's nursery-school teachers expose them, daily for five days, to a specially prepared book that paired nonwords featuring consonants typically learned early or late in Italian with pictures of 'fantasy animals'. Only infants showing consistent, stable use of one or more consonants gave evidence of recognizing the trained form-meaning pairs at test, and that only for the nonwords with consonants that are typically learned early.

These findings strongly suggest that vocal production practice provides the infant with the more robust word-form representations that are required for remembering specific object names, for first word production and for the learning of completely unfamiliar words. That is, the development, for an often-repeated vocal form, of a motor representation of that form (and the associated auditory representation) means that hearing related word forms (or attending to familiar objects with related labels) can prime the relevant form, sometimes resulting in identifiable early word use; those motor/auditory representations can also facilitate the mapping and retention of novel word forms and their meanings.

However, good evidence of a specific production effect on memory for individual words (or nonwords) has been hard to come by. Some experimental studies have shown such an effect, in both adults (Zamuner et al., 2016) and five-year-olds (Icht & Mama, 2015). Others find no effect (Hodges, Munro, Baker, McGregor, Docking & Arciuli, 2016) or even a reverse effect (in adults: Leach & Samuel, 2007; in toddlers: Zamuner et al., 2018), suggesting that producing a new word form immediately after first exposure may block learning; the attentional demands may be too great. Assef, Desmeules-Trudel, Bernard & Zamuner (2021) provide some clarification of these contradictory effects, reporting a developmental shift in the production effect between 2 and 6 years: For the youngest children, but not for the older ones, *saying* a new word (as compared with merely hearing it) made it less likely that they would retain it for recognition a little later (although there were considerable individual differences at all ages; a study based on vocabulary size rather than age might clarify further). These findings suggest that, in the earliest period of lexical development, what best supports word learning may be an increase in the number of phonological patterns or sequences that have become familiar from production (essentially, an increase in expressive vocabulary size) rather than immediate production or imitation of the novel word itself. The findings are consistent with studies showing clear lexicality effects on NWR, such as Dollaghan et al. (1995) and Cychosz et al. (2021), as well as with studies showing that expressive vocabulary strongly predicts both lexical processing efficiency and new word learning (e.g., Hodges et al., 2016; Torkildsen, Svangstu, Hansen, Smith, Simonsen, Moen & Lindgren, 2008).

The message that we can take from these efforts to experimentally determine the effect of production on word learning is that the effect, at least for the early word learning period, likely derives from a global benefit of lexical knowledge based on use, not from specific experience articulating a given word. Furthermore, articulatory skill per se has proven not to be a good predictor of word learning (Hodges et al., 2016): Relative skill in producing speech sounds is less important for retaining newly encountered words than experience with words more generally (overall vocabulary size). In short, early word learning is best facilitated by previous experience with speech forms overall, especially with one's own output. This conclusion is supported by the fact that although attempts at experimental capture of a specific effect of production on memory for just-heard forms have generally been unsuccessful with the youngest children, Keren-Portnoy et al. (2010) showed that *overall familiarity* (from production) with subcomponents (syllables or segments) of a novel form leads to greater accuracy in NWR. Importantly, where key component sounds were unfamiliar, the errors children made involved not only those sounds but also *other aspects* of the word forms that included them, reflecting the holistic nature of children's word-form representations. (See Dollaghan et al., 1995, for a related finding with older children.)

The origins of phonological memory: Vocal production and first words

To demonstrate how phonological memory may develop in infancy I turn now to evidence regarding the course of early word learning. I distinguish two phases: (i) the onset of established word use, which has been identified in observational studies as the spontaneous use of four words in a half-hour session recorded in the home with a familiar caretaker and a familiar visiting researcher (four word point, or 4wp) and (ii) the end of the single-word period, identified as the first session in which 25 or more words are produced spontaneously (25wp; word combinations are generally observed in the same session or within the month that follows). It is important, in this early period of word use, to compare children's phonological advances based on lexical milestones, not age, given the considerable range in learning rates. For example, of 59 British children recorded weekly from 9 months to the point of having two VMS and then monthly to 18 months (DePaolis et al., 2016), only six reached the 4wp by 12 months, another 23 by 15 months and all but four by 18 months; only 16 had reached the 25wp before the study ended.

The first phase can be characterised as 'item learning': There is generally little good evidence of systematicity (that is, of the different words being produced in similar ways, with a single prosodic structure, for example, or the same onset consonant: Menn & Vihman, 2011). Table 1 provides the first words recorded for one child each in six languages, based on recordings of caretaker-child free play made in each case by a native-speaker observer, who then transcribed the child's vocalizations and followed an established procedure to identify any words produced (Vihman & McCune, 1994). The children are ordered by age at 4wp.

Table 1. Words from the first recorded session with at least four spontaneously produced words. Stress on first syllable unless marked. Data are from Vihman, 2019 unless otherwise indicated; ordering of data sets is by age at first recorded word production.

Italian: Anna, 10-13 mos. (Keren-Portnoy, Majorano & Vihman, 2009)
Note that these first words were produced over three sessions.

target word	***gloss***	***adult form***	***child form***
bambola	doll	/bambola/	[bombœ]
bebè	baby	/bebˈe/	[bebɛ]
caffè	coffee	/kaf:ˈe/	[kakɛ]
cavallo	horse	/kaˈval:o/	[kak:o]
dindon	ding- dong	/dindon/	[dende]
mamma	mama	/mam:a/	[mam:a]

French: Charles, 11 mos.

target word	***gloss***	***adult form***	***child form***
au-revoir	byebye	/ɔʁvwaʁ/, /ɔvwaʁ/	[awa, haɥa]
boum	boom	/bum/	[ba, bœm]
beau	beautiful, good	/bo/	[bo]
donne, tiens	give/here	/dʌn/, /tjɛ̃/	[dæ]
mama	mama	/mama/	[mama]
non	no	/nɔ̃/	[nɛ]

Mandarin: Xinyu, 11 mos. (Lou, 2021)
The numbers following the segmental transcriptions indicate tones: 1 high level, 2 rising, 3 falling-rising, 4 falling. 0 is the neutral tone, which occurs only in unstressed syllables; it is shorter than the others and more phonetically compressed. (See Lou, 2021.)

target word	***gloss***	***child form***
/miau1/	meow	[mɛ1], [mi4], [miɛ1me4], [miɛ1mi4]
/pai1pai1/	byebye	[pa1pa0], [pɛ1pɛ4], [pa4pa4]
/waŋ1/	woof	[wa1]
/wei2/	hello (telephone)	[wei2], [wei4], [wa4], [wa1wa0]
/wo3/	me	[wo4], [wo4wo4]

Finnish: Eliisa, 13 mos.

target word	***gloss***	***adult form***	***child form***
hau(hau)	woof	/hɑu/	[hɑʊ], [hɑ], [hɑʊhɑʊ]
kiikkuu	is swinging	/ki:k:u:/	[ki:k:ʊ̥], [ki:ko], [ki:], [ki:kao]
lamppu	lamp	/lɑmp:u/	[pɑpʊ]
pallo	ball	/pɑl:o/	[pɑpʊ̥], [pɑpɑ̥u̥], [bɑpa], [pɑpʊ], [pɑp:ʊ], [pʌpɑ̥]

Welsh: Fflur, 13 mos.

target word	***gloss***	***adult form***	***child form***
agor	*open*	/agɑr/	[gəga], [gəg:a:], [gɒg:a], [gʌg:a]
blodyn	*flower*	/blɒdɪn/	[ʔɛbə], [bəba], [bəʊwa], [bəb:ə]
ceg	*mouth*	/kɛg/	[gag], [gɛ:g]
golau	*lights*	/gɒla/	[baʊwa], [ʔʌwa]
na	*no*	/na/	[na(:)], [ɲa]:, [ɹa::], [naʔ], [næ:]
sannau	*socks*	/sana/	[dənəʰ]
sgidie	*shoes*	/sgidijɛ/, /sgidʒɛ/	[gəga], [dadæ], [dada]

British English: Jude, 14 mos.

target word	***adult form***	***child form***
ball	/bɒl/	[baba]
barnaby	/ˈbɑnəbi/	[babi]
caterpillar	/ˈkætəpilə/	[biɔ]
yeah	/jɛə/	[jə]

Ferguson and Farwell (1975) were the first to call attention to the surprising finding that first identifiable word forms are, on the whole, relatively 'accurate', or phonetically close to their adult targets – and that the targets themselves tend to be quite simple (typically no more than one or two syllables in length and often with only a single consonant type per word). The relative accuracy is evident in Table 1, where the child words contain only consonants derived from the target, in the target order (or with typical child substitutions, such as stop for fricative or glide for liquid) – although three children (learning Finnish, Italian or Welsh) simplify their targets by creating harmony forms: This affects about half of their first words. (For first-word data from 48 children representing 10 languages, see Menn & Vihman, 2011.) There is little if any sign of reordering of syllables or segments. As to the adult forms, the word length constraint is seen in all but four (two each from English and Italian) and codas are relatively rare, even in a language like English, in which most content words have codas (Vihman et al., 1994).

Ferguson and Farwell drew from their findings the radical conclusion that first words may be pre-selected on phonological grounds, which seemed to suggest precocious infant 'awareness' of what they are unable to produce (Menn, 1983). Vihman (1996) proposed an alternative hypothesis, namely, that rather than avoiding difficult targets, infants implicitly filter input speech through a motorically-based response to frequently occurring word forms that roughly match their own highly familiar vocal output.

> 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. (p. 142)

Empirically, this idea is supported by the finding that the first words of a given child tend to be quite similar to the forms that child is using in babble (Vihman, Macken, Miller, Simmons & Miller, 1985). Theoretically, it also draws on the fact that the cross-modal experience involved in vocal production (proprioceptive experience as well as perception of the concomitant auditory effects) robustly familiarizes the child with his or her own most-used vocal patterns; this could be expected to lead to a pop-out effect for similar adult forms in input speech, rendering those forms likely to be retained along with their linked articulatory routine.

The hypothesis gains broader theoretical support from Thelen and Smith's (1994) account of the strengthening of neural connections from concurrent experiences in the action and perception systems. 'We argue that it is this temporal association of perception and action that lays the foundation for behavioral development' (p. 142). Elaborating on this, Thelen and Smith note that Edelman, in his Theory of Neuronal Group Selection (e.g., 1987),

> emphasizes selection as the primary developmental mechanism: while epigenetic processes produce brain structures that are modally similar in all members of a species,

> these same processes also produce – through selection – enormous individual variability at the level of anatomical connectivity of the individual neurons and groups…*It is the temporal coherence of sensory and motor signals that acts as the selective process.* (p. 142f. [italics added])

The proposed articulatory filter encapsulates this idea with respect to the first child word production. Three of the experimental studies cited above (DePaolis et al., 2011, 2013; Majorano et al. 2014) were designed to test the articulatory filter hypothesis. Although, as noted, the results were more complex than expected, with infants who met criteria for more than one VMS unexpectedly showing a *novelty* response, they nevertheless provide solid support for the idea that vocal production affects the infant's processing of input speech (see now also Vilain, Dole, Loevenbruck, Pascalis & Schwartz, 2019, who demonstrate an effect of infant vocal production on early consonant category formation). The results suggest that infants who have demonstrated mastery of only a single supraglottal consonant are attracted more to familiar sounds, filtering the stimuli as predicted by the hypothesis, while those who have begun to build a stable repertoire of two or more consonantal production routines shift their attention to what remains unfamiliar. In line with this, the 10-month-olds who showed the novelty response to untrained word forms in DePaolis et al. (2016) were also more lexically advanced, based on CDI reports, than either the children who showed the familiarity response or those who failed to show any preference at all.

The suggested pathway from vocal practice to first words, then, is twofold: (i) Babbling experience lends salience to some repeatedly encountered sequences in the input; (ii) the production-based salience of those selected patterns provides automatic access to the associated motor routines, which can lead to the situational priming of first words (Vihman, 2017). However, once a child has gained experience with production of more than a single VMS, sound patterns not yet in repertoire elicit a greater response, reflecting, perhaps, a kind of 'habituation' to the familiar and an awakened interest in what is new (Hunter & Ames, 1988). The child's repeated use of their own patterns in early vocalizations and first word production strengthens the familiarity of those patterns. In effect, those patterns constitute an emergent 'phonological memory', not a time-limited temporary store or window for recall but a top-down filter that automatically activates familiar vocal forms on hearing similar elements in the ongoing flow of speech; this supports retention of selected patterns and thereby facilitates new word learning (or, where required, nonword repetition).

Furthermore, as the initial sample or database of self-produced word forms accumulates, a two-part attentional response can again be expected. The small set of word forms that have come to be known from production will have gained proprioceptive in addition to auditory familiarity (as well as a link to situations of use that strengthens with every new exposure: McMurray, Horst & Samuelson, 2012). At first, hearing one of those newly established words will automatically activate both a motor representation and the cloud of related auditory exemplars, with each such repeated experience leading to better retention of the word and a bias toward continued targeting of similar word forms. A matter of days or weeks later – with the timing differing by the individual child's willingness to depart from their 'comfort zone' to explore novel patterns – a shift to more challenging, less familiar forms will occur, initiating the next phase in the development of word production.

Expansion of the lexicon and origins of a phonological network

Once word use is established, children begin producing new words more regularly. For phonological development, the most striking aspect of this next phase is a *decrease* in accuracy or fidelity to the target form: Children now begin attempting more challenging word forms, which may include prosodic structures or phonetic elements with which they lack sufficient articulatory skill or experience for accurate production. What we typically find is that rather than somehow rapidly increasing their phonetic skills to accommodate the new forms, children begin to extend their production routines, adapting or assimilating target words to their existing motoric repertoire, arriving at individual ‘templates’ or favored production patterns that do not always respect the presence or sequencing of elements of the target form. Table 2 illustrates the adaptation of challenging forms to a pattern familiar to the child, using 25wp data from the children whose first words are shown in Table 1. In each case some word forms (marked with a star) are ‘adapted’ to fit a particular pattern (given in angle brackets for each child), while other words the child is producing can be considered ‘selected’ (i.e., their target form already fits the pattern). Whereas Table 1 presents all the child’s first recorded words, Table 2 picks out for illustration one pattern or template for each child but presents all the words recorded in the 25wp session that fit that pattern.

Table 2. Words from the 25-word-point session, chosen to illustrate one production pattern for each child. Stress is on first syllable unless marked. * marks words adapted to fit pattern. < > indicates pattern or template; subscript $_0$ indicates an optional element. Data are from Vihman (2019) unless otherwise indicated; ordering is by age at 25wp. The children are ordered by age at 25wp.

British English: Jude, 15 mos. <C_0VGLIDEV>

target word	***adult form***	***child form***
*babi ‘baby’[a]	/bɑbɨ/	[bawa]
*caterpillar	/kætəpɪlə/	[pi[j]a]
hello	/hɛ'ləʊ/	[aijɔ]
hiya	/hɑɪjʌ/	[aija], [jaija]
*in (th)ere	/ɪne[j]ə/	[ɪjɛ]
*orange	/ɔrənʤ/	[ɒwi]
*yeah	/jeə/	[wəjɛ]

[a]Welsh word (recorded in N. Wales)

Finnish: Eliisa, 15 mos. <C_1VC_1V> (consonant harmony)

target word	***gloss***	***adult form***	***child form***
kaka	caca	/kɑkɑ/	[kaka]
kakku	cake	/kɑk:u/	[kak:o]
kiikkuu	is swinging	/ki:k:u:/	[ki:k:u]
kukka	flower	/kuk:ɑ/	[kuk:a]
kukkia	some flowers	/kuk:iɑ /	[kuk:ia]
*lintu	bird	/lintu/	[titu], [tito]
mummu, muumuu	grandma	/mum:u/, /mu:mu:/	[mum:u], [mu:mu:]
*nuke	doll	/nuke/	[kuk:e]

*nukkuu	is sleeping	/nuk:u:/	[kuk:u:]
*pallo	ball	/pɑl:o/	[pap:u]
*paperi	paper	/pɑperi/	[papa], [pape]
pappa	grandpa	/pɑp:ɑ/	[pap:a]
pomppi	jump!	/pomp:i/	[pop:i]
pupu	bunny	/pupu/	[pupo], [pop:i:]
*rikki	broken	/rik:i/	[kik:i]
*tippu, tippuu	drop, drops	/tip:u/, /tip:u:/	[pip:u]
tyttö	girl	/tüt:ö/	[tit:ɔ]

Welsh: Flur, 17 mos. <GLOTTALVCVC$_0$>

target word	***gloss***	***adult form***	***child form***
*bwni	bunny	/buni/	[ʔɛɲ:ɛ], [hʊni]
eto	again	/eto /	[ʔəθax]
*fyna	over there	/vɨnɑ/	[ʔɛn:æ:], [ʔəm:a]
hwnna	this	/hunɑ/	[henɛ], [ʔəɲa], [ʔɔn:a]
*moron	carrot	/moron/	[ʔɒwan], [həwɛn]
*Po	TV character	/po/	[hɔa:]
*sos goch	red sauce	/sosgox/	[ɔʔɔx]
*syrthio	fall	/sɨrθiɔ/	[ʔʌs:a]

French: Charles, 18 mos. <(h)aCa>, <(h)aCMIDV>

target word	***gloss***	***adult form***	***child form***
allo	hello (telephone)	/alo/	[al:ɔ]
assis	sitting	/asi/	[asa]
attend	wait	/atã/	[atæ]
au revoir	goodbye	/ɔʁvwaʁ/, /ɔvwaʁ/	[avwa]
auto	car	/oto/	[a::to]
*c'est beau	it's beautiful	/sebo/	[habo]
*c'est bon	it's good	/sebõ/	[habõ]
*chapeau	hat	/ʃapo/	[a:pø]
*garçon	boy	/gaʁsõ/	[aʐa[, [haʒœ]
*lapin	rabbit	/lapɛ̃/	[apa], [haba]
*non-non	no-no	/nõnõ/	[hənɔ̃]
*les yeux	eyes	/lezjø/	[aʒo:]
*oiseau	bird	/wazo/	[abo:], apo]
*va pas	doesn't go	/vapa/	[apa], [hapa]
*voire	to see	/vwaʁ/	[abvwa], [av:wa]

Mandarin: Xinyu, 17 mos. (Lou, 2021) <$C_1V_1(V_2)C_1V_1(V_2)$> (reduplication)

The numbers following the segmental transcriptions indicate tones: 1 high level, 2 rising, 3 falling-rising, 4 falling. O is the neutral tone, which occurs in unstressed syllables; it is shorter than the others and more compressed, phonetically. See Lou, 2021.

target word	***gloss***	***child form***
/jiɛ2jiɛ0/	grandpa	[jiɛ1jiɛ0]
/kɤ1kɤ0/	elder brother	[kɤ1kɤ0]
/la1la0/	TV character	[la1la0], [la1la4]]
/ma1ma0/	mummy	[ma1ma0]
/mei4mei0/	younger sister	[mei4mei0]
/mau1mau0/	kitten	[mau1mau1]
/nai3nai0/	grandma	[nɛ1nɛ0], [nɛ1nɛ1]
/pa4pa0/	daddy	[pa4pa0]
/pai1pai2/	bye-bye	[pa4pa0], [pa1pa0]
/pei1pei1/	cup	[pei1pei2], [pei1pei0]
*/piŋ3kɛn1/	biscuits	[ka1ka0], [pei4pei0], [pei1pei1]
/pʰa4pʰa4/	sound of clapping	[pʰa4pʰa0]
/ti4ti0/	younger brother	[ti4ti0]
/tu4tu0	bunny	[ty1tɛ0]
*/ʈʂɤ4kɤ0/	this	[ku4ku0]

Italian: Anna, 20 mos. (Keren-Portnoy et al., 2009) <$(CV)_0$CVlV>

target word	***gloss***	***adult form***	***child form***
*animali	animals	/aniˈmali/	[mali]
bambola	doll	/bambola/	[bambala]
*cappello	hat	/kaˈp:el:o/	[pɛl:o]
*cavallo	horse	/kaˈval:o/	[kalol:o]
gallo	rooster	/gal:o/	[gal:o]
giallo	yellow	/dʒal:o/	[dʒal:o]
*maiale	pig	/maˈjale/	[male]
*piccolo	little	/pik:olo/	[pilil:o]

The data presented in Table 2 reflect some differences between the ambient languages. Mandarin, for example, features a good deal of reduplication in input speech, likely because reduplication is used to express diminutive or affective meaning; our data show how little accommodation is needed for a child to match those patterns (though the children's use of tone is not yet generally target-like). In contrast, we see a surprising lack of fidelity to the word-initial consonant in the data of the children exposed to French or Welsh. This can be accounted for by target language prosody: French accents the last syllable of a phrase (but lacks lexical stress per se), while Welsh lengthens the final consonant of a stressed syllable, which is typically the first in disyllabic words; both of these rhythmic characteristics tend to deflect attention from the word-initial consonant (Vihman & Croft, 2007; see Vihman & Majorano, 2017, for experimental evidence of a similar perceptual effect of medial geminates in Italian).

At the same time, however, even children acquiring the same language show a wide range of individual differences at this point in their development (Vihman et al., 1994; Vihman, 2019). The data shown here are not meant to be 'typical' of the languages in question but only to show how relatively simple child production patterns have now begun to be extended to less accessible adult word forms, with each child settling on one or more patterns that have become familiar through the preceding months of babble practice and early word use and that thus constitute production routines readily available to assimilate novel forms. What is apparent from this selection of child word forms is that each child repeatedly uses similar structures or segments, so that many words that are not similar in their adult form come to sound alike in the child's form; this is the hallmark of emergent phonological systematicity. At this point the words the child produces are no longer isolated bits of knowledge; they have become linked with other forms that share a structure or an onset or coda segment, for example. That is, the child's extension of similar patterns to novel and challenging word forms reflects the emergence of a lexical network, a forging of links across known words, which can be taken to be the developmental equivalent of the 'lexical engagement' that Leach and Samuel (2007) identified in experimental work with adults learning novel word forms (see also the similar concept of 'lexicalization' in studies by Gaskell and colleagues, e.g., Gaskell & Dumay, 2003). The extension of a familiar production routine to new words is good evidence for the implicit function of phonological memory, namely, support from familiar patterns for the retention of (roughly similar) new forms.

An alternative account of phonological memory

Gathercole's (2006) account of phonological memory, like her earlier collaborative studies with Baddeley and others, emphasizes the central importance of a 'phonological store'*:*

> Repetition of nonwords necessarily requires the storage of its constituent phonological segments in the short-term store, and…the quality of this storage varies markedly between individuals… (p. 520).
>
> Nonword repetition ability is significantly constrained by phonological storage capacity, and…this capacity plays a key role in supporting learning of the sound structure of new words during vocabulary acquisition…Initial encounters with the phonological forms of novel words are represented in the short-term store, and…these representations form the basis for the gradual process of abstracting a stable specification of the sound structure across repeated presentations…[This] is a primitive learning mechanism that is particularly important in the early stages of acquiring a language. (p. 521)

The presumption that word learning is rooted in phonological representations that must first be retained in a time-limited store has tended to be accepted even by developmentalists who hold alternative views as to the nature of the problem for children who perform poorly on NWR (e.g., Edwards & Lahey, 1998; Bowey, 2006). However, this theoretical conceptualization has long been challenged by Macken and Jones and their colleagues, based on experimental short-term memory studies with adults (recent studies include Macken, Taylor, Kozlov, Hughes & Jones, 2016, Jones & Macken, 2018, G. Jones, Justice, Cabiddu, Lee, Jao, Harrison & Macken, 2020). These authors argue for an embodied conceptualization of the sources of what is usually interpreted as differences in 'phonological memory capacity'; their model involves no 'store' (Macken & Jones, 2003) but only the mutual activation and interaction of domain-general auditory perception and motor control processes. In this model, the construct ‘phonological memory’ has no independent nominal status but refers to the transitory product of phonological processing, in which patterns or sub-patterns familiar from previous perceptual and motoric experience guide or filter the brain’s response to ongoing speech input. This sensorimotor approach, grounded in brain studies as well as in behavioural evidence, readily lends itself to the developmental account of the origins of phonological memory in babble and early word production that I have laid out here.

Munson (2006) made an important point that has too often been disregarded in studies of phonological memory – namely, that the *segmental sequence* obtained as evidence of what the child can remember in the NWR task is actually an artefact of the analysis:

> The level of detail that [phonetic transcriptions] code is ultimately related more to the perceptual abilities of the listener, the degrees of freedom in the symbol system, and a priori assumptions about the quantity of detail that is relevant for transcription than to the signal being transcribed and its associated phonological representation. (p. 578)

(The point is necessarily also relevant with regards to the data presented above – but this is in any case consistent with our claim that the child’s representation is holistic, at least in part, rather than fully specified segmentally.)

D. M. Jones and Macken (2018) make a similar point, with reference to adult verbal short-term memory (vSTM) studies, in their forceful rejection of Baddeley’s cognitive model: ‘We view the concepts of vSTM and its concomitant processes as reifications from observations of performance on these tasks’ (p. 351). They explain that the ‘dynamics and gradience’ of the (oral) response provided by participants in such tasks are obscured by the way in which the response is recorded, because it

> is coded in terms of discrete verbal entities. These are then compared with the presented sequence and evaluated accordingly to infer what happened to the sequence’s content within the metaphoric store… This evaluation involves only comparison of two ordered sequences of discrete entities... (p. 354)

In other words, both Munson (2006) and Jones and Macken (2018) identify a fundamental flaw in the box-and-arrow cognitive model, namely, its dependence on an interpretation of the written encoding of the participant’s response as literally reflecting the production of a sequence of segments rather than a co-articulated flow of speech sounds, further enriched by their prosodic envelope. (For evidence of the role of co-articulation in supporting recall of sequences of items in adult vSTM studies, see Woodward, Macken & Jones, 2008.)

Port & Leary (2005) note more broadly, with regards to ‘formal phonology’, that ‘there is no evidence…for a memory [for segments or phonological categories] that is exclusively serial’ (p. 933). Arguing for a new approach to phonology altogether, Port and Leary make the strong claim that ‘there is no discrete universal phonetic inventory and thus phonology is not amenable to formal description’ (p. 952). Among the consequences that they draw from this the most pertinent for our purposes is that ‘speakers can record in memory and control in their production far more detail than traditional linguistics supposed’ (p. 953). The authors suggest that ‘phonological categories…may have both an abstract representation and also something like a cloud of specific examples or episodes of concretely specified events [i.e., exemplars]’ (p 953).

Port and Leary mention language development only in passing, but their approach is compatible with the account of holistic representations that Vihman and Keren-Portnoy (2013) provide, based on observation of infants' production in the single-word period:

> The child's rendition of the word shows sensitivity to some of the segments or phonetic features that occur in it. However, it does not necessarily maintain the order in which the segments or features appear, but may instead redistribute, merge or spread some of those features…Within the lexical unit there may be no clear evidence that the child has registered information about the identity and number of all of the segments (as perceived by adult speakers) or their relative order. Based on evidence from production, then, children seem to have a representation or memory trace of the adult form, but that representation is not constructed out of an ordered sequence of segments.... In fact, as we understand them, these child representations include abundant detail – much more than is apparent in phonemic or even broad phonetic description. (p. 5f.)

G. Jones et al. (2020) address the contrasting interpretations of the effects of development on verbal STM based on corpus analysis of samples of the kind of speech children hear and with the use of computational modelling to support an associative learning approach. These authors warn against

> construing STM as a primitive processing system, [instead of] as a particular setting within which participants must deal with a relatively novel situation by flexibly co-opting whatever skills and knowledge they have to accomplish the particular task goals. (p. 2)

Going a step further, D. M. Jones and Macken (2018) see 'even '"[short-term] memory" and "forgetting"' as 'hypothetical constructs proposed to explain…behaviour' on vSTM tasks (p. 354). The dismissal of memory itself is perhaps best understood as suggesting that, within the somewhat artificial context of vSTM tasks, including NWR (with forms presented outside of any syntactic, semantic or pragmatic frame), what has been taken to be memory retrieval is better construed as verbal production primed by what has just been heard and further informed by whatever elements are familiar already – that is, by previous experience and the domain-general processes of associative learning. Consistent with this, G. Jones et al. relate responses to vSTM tasks to the basic sensorimotor (non-verbal) character of the processes of perceiving and acting, arguing that

> associative learning is … heavily embedded in familiarity – items and item sequences that are often encountered in the language environment will be more amenable to learning than those that are not. (p. 104200)

Finally, to the evidence of a role for existing knowledge in facilitating novel word production G. Jones et al. (2021) contribute the findings of a modelling study designed to test the idea that word learning is fundamentally based on a gradual build-up of familiar 'chunks of phoneme sequences' (p. 11). The model, a 'chunk-based learning mechanism' designed to gain knowledge of increasingly long phoneme sequences from repeated exposure to language data, showed effects of word frequency, word length, phonotactic probability and neighborhood density very similar to those that account for which words are learned earliest in a longitudinal corpus of transcribed child data (covering ages 2 to 3 years). Most importantly, the study demonstrates that

> the amount and type of knowledge already acquired at the point of learning has a major influence on how the child's input is processed, and thus how new knowledge is acquired and integrated into long-term memory. (p. 11).

This account is in accord with the developmental origins of phonological memory as I have outlined them here, as are the findings of Verhagen et al. (2019) and Kehoe et al. (in press), which suggest a particularly strong effect of existing knowledge on word learning in children under age five. None of these authors mention self-action, or familiarity with one's own production. However, I see action initiated by the self as central to an understanding of the transition into language, as it is to the Dynamic Systems Theory account of infant development more generally (Thelen & Smith, 1994; for discussion and illustration of how closely Dynamic Systems principles accord with this understanding of phonological development in the transition period, see Vihman, DePaolis & Keren-Portnoy, 2015; Vihman, 2019, ch. 1). These views also fit in well with embodied and associative learning approaches more generally.

To briefly recapitulate this account, babble practice and the consequent auditory feedback support the retention and recall of heard word patterns; production of these remembered forms in turn creates sensorimotor knowledge of a growing range of first shorter, then longer forms. The traces (or exemplars) of heard words and the phonological representations that emerge from those traces stabilize with repeated use and create the flexibility for the child to attempt more challenging words using the existing routine(s) (i.e., their preferred prosodic structures or templates). As more words are learned and phonetic skills continue to improve, phonological memory develops further, facilitating more word capture. Of course, this developmental story does not begin to take into account the parallel emergence of an understanding of reference and of a semantic network, advances which, like the process of building up phonological memory, will proceed at differing rates in different children. In Dynamic Systems terms, any one of these aspects of word learning may take the lead (or serve as a 'control parameter') in triggering a non-linear lexical and/or phonological phase shift in the individual child.

In this account, then, familiarity with one's own sound patterns provides a foothold for entering into the input speech stream and retaining selected bits; eventually, this leads to the production of interpretable word forms in priming contexts. Over time and with further exposure to the language, familiarity with a growing expressive database of syllables and word forms creates the foundation for ever more effective sensorimotor retention of novel lexical items and also supports the creation of an increasingly dense network of linked forms (and meanings). Vocal motor patterns can be seen as the source of phonological memory, which need not be imagined as a cognitive box or store of any kind. Instead, we can conceptualize it as referring to the unavoidable process of accessing what is familiar while experiencing what is new.

References

Assef, B. L., Desmeules-Trudel, F., Bernard, A. & Zamuner, T. S. (2021). A shift in the directio of the production effect in children aged 2-6 years. *Child Development*. doi: 10.1111/cdev.13618

Baddeley, A. (1986). *Working Memory*. Oxford: Clarendon.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Reiew., 105*, 158–173. doi:10.1037/0033-295X.105.1.158

Bergelson, E. & Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Science, 109*, 3253-3258. doi:10.1073/pnas.1113380109

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics, 22,* 441-469. doi: 10.1017/S0142716401003083.

Bowey, J. A. (2006). Clarifying the phonological processing account of nonword repetition. *Applied Psycholinguistics, 27,* 548-552. doi:10.1017.S0142716406060401

Boysson-Bardies, B., Hallé, P., Sagart, L. & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language, 16*, 1-17. doi: 10.1017/S0305000900013404doi:10.1017/S0305000900013404

Boysson-Bardies, B. de & Vihman, M. M. (1991). Adaptation to language. *Language, 67*, 297-319. doi:10.2307/415108

Brown, R. (1958). *Words and Things.* New York, NY: Free Press.

Conboy, B. T. & Mills, D. L. (2006). Two languages, one developing brain. *Developmental Science, 9,* F1-F12. doi:10.1111/j.1467-7687.2005.00453.x

Connine, C , Blasko, D., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition*? Journal of Memory and Language, 32*, 193-210. doi: 10.1006/jmla.1993.1011

Cychosz, M., Erskine, M., Munson, B. & Edwards, J. (2021). A lexical advantage in four-year-old children's word repetition. *Journal of Child Language, 48,* 31-54. doi: 10.1017/S0305000920000094

Davis, M. & Redford, M. A. (2019). The emergence of perceptual-motor units in a production model that assumes holistic speech plans. *Frontiers in Psychology, 10,* 2121/1-19. doi: 10.3389/fpsyg.2019.02121

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, 1-11. doi:10.3389/fpsyg.2016.00715

DePaolis, R., 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. doi:10.1080/15475441.2013.829740

DePaolis, R., Vihman, M. M. & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development, 36*, 642-649. doi:10.1016/j.infbeh.2013.06.007

Dollaghan, C. A., Biber, M. E. & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics, 16*, 211-222. doi:10.1017/S0142716400007098

Edelman, G. (1987). *Neural Darwinism*. New York: Basic Books.

Edwards, J. & Lahey, M. (1998). Nonword repetitions of children with specific language impairment. *Applied Psycholinguistics, 19*, 279–309. doi:10.1017/S0142716400010079

Fagan, M. K. (2009). Mean length of utterance before words and grammar. *Journal of Child Language, 36,* 495-527. doi:10.1017/S0305000908009070

Ferguson, C. A. & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51*, 419-439. doi:10.2307/412864

Fowler, A. E. (1991). How early phonological development might set the stage for phonological awareness. In S. Brady & D. Shankweiler (eds.), *Phonological Processes in Literacy* (pp. 97-117). Hillsdale, NJ: Erlbaum.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89,* 105–132. doi: 10.1016/S0010-0277(03)00070-2

Gathercole, S. E. (2006). Nonword repetition and word learning. *Applied Psycholinguistics, 27.*513-543. doi:10.1017.S0142716406060383

Gathercole, S. E. & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children. *Journal of Memory and Language, 28,* 200–213. doi:10.1016/0749-596X(89)90044-2

Gathercole, S.E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12,* 349-367. doi: 10.1017/S0142716400009267

Gathercole, S.E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years. *Developmental Psychology, 28,* 887-898. doi: 10.1037/0012-1649.28.5.88

Hallé, P. & Boysson-Bardies, B. de. (1994). Emergence of an early receptive lexicon. *Infant Behavior and Development, 17,* 119-129. doi:10.1016/0163-6383(94)90047-7

Hauser, M. D. (1996). *The Evolution of Communication.* Cambridge, MA: MIT Press.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetic, 31*, 373–405. doi:10.1016/j.wocn.2003.09.006

Hodges, R., Munro, N., Baker, E, McGregor, K., Docking, K. & Arciuli, J. (2016). The role of elicited verbal imitation in toddlers' word learning. *Journal of Child Language, 43*, 457-471. doi:10.1017/S0305000915000240

Hoff, E., Core, C. & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20-to 24-month-olds. *Journal of Child Language, 35*, 903-916. doi:10.1017/S0305000908008751

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.

Icht, M. & Mama, Y. (2015). The production effect in memory. *Journal of Child Language, 42,* 1102-1124. doi:10.1017/S0305000914000713

Johnson, K. A. (2007). Decisions and mechanisms in exemplar-based phonology. In M-J. Solé, P. S. Beddor & M. Ohala (eds.), *Experimental Approaches to Phonology,* pp. 25-40. Oxford: OUP.

Jones, D. M. & Macken, B. (2018). In the beginning was the deed. *Current Directions in Psychological Science, 27,* 351-356. doi:10.1177/09637214187657

Jones, D. M., Macken, W. J. & Nicholls, A. P. (2004). The phonological store of working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 656–674. doi:10.1037/0278-7393.30.3.656

Jones, G., Justice, L. V., Cabiddu, F., Lee, B. J., Iao, L-S., Harrison, N. & Macken, B. (2020). Does short-term memory develop? *Cognition, 198, 104200* doi:10.1016/j.cognition.2020.104200

Jones, G., Cabiddu, F., Andrews, M. & Rowland, C. (2021). Chunks of phonological knowledge play a significant role in children's word learning and explain effects of neighborhood size, phonotactic probability, word frequency and word length. *Journal of Memory and Language, 119,* 104232. doi.org/10.1016/j.jml.2021.104232

Jones, G. & Macken, B. (2018). Long-term associative learning predicts verbal short-term memory performance. *Memory & Cognition, 46,* 216-229. doi.org/10.1177/09637214187657

Jusczyk, P. W. (1997). *The Discovery of Spoken Language.* Cambridge, MA: MIT Press.

Kehoe, M., Poulin-Dubois, D. & Friend, M. (in press). Within- and cross-language relations between phonological memory, vocabulary, and grammar in bilingual children. *Journal of Speech, Language and Hearing Research.*

Kent, R. D. & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Socieety of America, 72*, 353-363.

Keren-Portnoy, T., Majorano, M. & Vihman, M. M. (2009). From phonetics to phonology. *Journal of Child Language, 36,* 235-267. doi:10.1017/S0305000908008933

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. doi:10.1044/1092-4388(2009/09-0003)

Laing, C. & Bergelson, E. (2020). From babble to words: Infants' early productions match words and objects in their environment. *Cognitive Psychology, 122*, 101308. doi:10.1016/j.cogpsych.2020.101308

Leach, L. & Samuel, A. G. (2007). Lexical configuration and lexical engagement. *Cognitive Psychology, 55,* 306–353. doi:10.1016/j.cogpsych.2007.01.001

Lee, C-C., Jhang, Y., Chen, L., Relyea, B. & Oller, D. K. (2017). Subtlety of ambient-language effects in babbling. *Language Learning and Development, 13,* 100-126. doi:10.1080/15475441.2016.1180983

Lou, S. (2021) Early phonological development in Mandarin. Unpublished PhD thesis, University of York.

Macken, B., Taylor, J. C., Kozlov, M. D., Hughes, R. W. & Jones, D. (2016). Memory as embodiment. *Cognition, 155,* 113–124. doi:10.1016/j.cognition.2016.06.013

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. New York: Academic Press.

Macken, W. J. & Jones, D. M. (2003). Reification of phonological storage. *Quarterly Journal of Experimental Psychology, 56*, 1279–1288. doi:10.1080/02724980245000052

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*, 499-546. doi:10.1017/S0140525X98001265

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. doi:10.1017/S0305000918000521

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. doi:10.1080/15475441.2013.829740

Marchman, V., Fernald, A. & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish–English bilinguals. *Journal of Child Language, 37,* 817-840. doi:10.1017/S0305000909990055

McCune, L. & Vihman, M. M. (2001). Early phonetic and lexical development. *Journal of Speech, Language and Hearing Research, 44*, 670-684. doi:10.1044/1092-4388(2001/054)

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? *Child Development, 88,* 156-166. doi:10.1111/cdev.12671

McMurray, B., Horst, J. & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119*, 831–877. doi: 10.1037/a0029872

Melby-Lervåg, M., Lervåg, A., Lyster, S. A., Klem, M., Hagtvet, B., Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science, 23,* 1092–1098. doi.org/10.1177/0956797612443833

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. In B. Butterworth (ed.), *Language Production*, vol. 2. London: Academic Press.

Menn, L., Schmidt, E. & Nicholas, B. (2013). Challenges to theories, charges to a model. In M. M. Vihman & T. Keren-Portnoy (eds.), *The Emergence of Phonology.* Cambridge: CUP.

Menn, L. & Vihman, M. M. (2011). Features in child phonology. In N. Clements & R. Ridouane (eds.), *Where Do Phonological Features Come From?,* pp. 261-301. Amsterdam: Benjamins.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3–19. doi:10.1037/0022-0663.91.1.3

Munson, B. (2006). Nonword repetition and levels of abstraction in phonological knowledge. *Applied Psycholinguistics, 27,* 548-552. doi:10.1017.S0142716406060474

Munson, B., Edwards, J. & Beckman, M. E. (2012). Phonological representation in language acquisition: Climbing the ladder of abstraction. In A. C. Cohn, C. Fougeron & M. K. Huffman (eds.), *Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press.

Noiray, A., Popescu, A., Killmer, H., Rubertus, E., Krüger, S & Hintermeier, L. (2019). Spoken language development and the challenge of skill integration. *Frontiers in Psychology, 10,* 2777. doi: 10.3389/fpsyg.2019.0277

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum.

Parra, M., Hoff, E. & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology, 108*, 113-125. doi:10.1016/j.jecp.2010.07.011

Pierrehumbert, J. (2016). Phonological representation. *Annual Review of Linguistics, 2,* 33-52. doi:10.1146/annurev-linguist-030514-125050

Poeppel, D. & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews, 21*, 322-334. doi:10.1038/s41583-020-0304-4

Port, R. F. & Leary, A. P. (2005). Against formal phonology. *Language, 82, 927*-964. doi: 10.1353/lan.2005.0195

Roy, P. & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year olds. *Journal of Speech, Language and Hearing Research, 47*, 223–234. doi:10.1044/1092-4388(2004/019)

Rvachew, S., Alhaidary, A., Mattock, K. & Polka, L. (2008). Emergence of the corner vowels in the babble produced by infants exposed to Canadian English or Canadian French. *Journal of Phonetics*, 36, 564–577. doi:10.1016/j.wocn.2008.02.001

Segal, O., Keren-Portnoy, T. & Vihman, M. M. (2020). Robust effects of stress on early lexical representation. *Infancy, 25*, 500-521. doi:10.1111/infa.12340

Snowling, M., Chiat, S. & Hulme, C. (1991). Words, nonwords, and phonological processes. *Applied Psycholinguistics, 12*, 369-373. doi:10.1017/S0142716400009279

Swingley, D. (2005). Eleven-month-olds' knowledge of how familiar words sound. *Developmental Science, 8*, 432-443. doi:10.1111/j.1467-7687.2005.00432.x

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society, B, 364,* 3617–3632. doi:10.1098/rstb.2009.0107

Thelen, E. & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Thierry, G. & Vihman, M. M. (2008). The onset of word form recognition. In A. D. Friederici & G. Thierry (eds.), *Early Language Development,* pp. 115-135. Amsterdam: Benjamins.

Torkildsen, J. V. K., Svangstu, J. M., Hansen, H. F., Smith, L., Simonsen, H. G., Moen, I. & Lindgren, M. (2008). Productive vocabulary size predicts Event-Related Potential correlates of fast mapping in 20-month-olds. *Journal of Cognitive Neuroscience, 20*, 1266-1282. doi:10.1162/jocn.2008.20087

Verhagen, J., Boom, J., Mulder, H., de Bree, E. & Leseman, P. (2019). Reciprocal relationships between nonword repetition and vocabulary during the preschool years. *Developmental Psychology, 55*, 1125-1137. doi:10.1037/dev0000702.supp

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, *21*, 61-82. doi: 10.1016/S0095-4470(19)31321-X

Vihman, M. M. (1996). *Phonological Development*. Oxford: Basil Blackwell.

Vihman, M. M. (2002). The role of mirror neurons in the ontogeny of speech. In M. Stamenov & V. Gallese (eds.), *Mirror Neurons and the Evolution of Brain and Language* (pp. 305-314). Amsterdam: John Benjamins.

Vihman, M. M. (2017). Learning words and learning sounds. *British Journal of Psychology, 108,* 1-27. doi:10.1111/bjop.12207

Vihman, M. M. (2019). *Phonological Templates in Development.* Oxford: OUP.

Vihman, M. M. & Boysson-Bardies, B. de. (1994). The nature and origins of ambient language influence on infant vocal production and early words. *Phonetica, 51,* 159-169.

Vihman, M. M. & Croft, W. (2007). Phonological development: Toward a 'radical' templatic phonology. *Linguistics, 45,* 683-725.

Vihman, M. M., DePaolis, R. A. & Keren-Portnoy, T. (2015). A Dynamic Systems approach to babbling and words. In E. L. Bavin & L. R. Naigles (eds.), *The Cambridge Handbook of Child Language,* 2nd ed, pp. 207-229. Cambridge: Cambridge University Press.

Vihman, M. M., Kay, E., Boysson-Bardies, B. de, Durand, C. & Sundberg, U. (1994). External sources of individual differences? *Developmental Psychology, 30*, 652-663. doi:10.1037/0012-1649.30.5.651

Vihman, M. M. & Keren-Portnoy, T. (2013). Introduction *The Emergence of Phonology*, pp. -14. Cambridge: Cambridge University Press.

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. DOI: 10.2307/414151

Vihman, M. M. & Majorano, M. (2017). The role of geminates in infants' early words and word-form recognition. *Journal of Child Language, 44,* 158-184. doi:10.1017/S0305000915000793

Vihman, M. M. & McCune L. (1994). When is a word a word*? Journal of Child Language, 21,* 517-542. doi:10.1017/S0305000900009442

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. doi:10.1016/j.jml.2003.11.004

Vihman, M. M., Ota, M., Keren-Portnoy, T. & Lou, S. (under review). A challenge to whole-word phonology? A study of Japanese and Mandarin.

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. doi: 10.1111/desc.12830

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review, 13*, 286-350.

Westermann, G. & Miranda, E. R. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language, 89,* 393-400. doi:10.1016/S0093-934X(03)00345-6

Woodward, A. J., Macken, W. J. & Jones, D. M. (2008). Linguistic familiarity in short-term memory. *Journal of Memory and Language, 58*, 48-65. doi:10.1016/j.jml.2007.07.002

Zamuner, T. S., Morin-Lessard, E., Strahm, S. & Page, M. P. A. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language, 89*, 55-67. doi:10.1016/j.jml.2015.10.003

Zamuner, T. S., Morin-Lessard, E., Strahm, S. & Page, M. (2017). Reverse production effect. *Developmental Science,* e12636. doi:10.1111/desc.12636