Learning words and learning sounds: Advances in language development

Marilyn M. Vihman

[1]University of York

Acknowledgement: Thanks are due to my colleagues Paul Foulkes and Amanda Cardoso, who provided useful feedback on a draft of the paper, to Rory DePaolis, who created the figure to illustrate two experiments, and to audiences at the ESPP conference in Tartu, Estonia, and the Child Language Seminar in Warwick, England, where versions of this paper were presented in the summer of 2015.

Word count (exc. figures/tables): 9446

*Requests for reprints should be addressed to Marilyn Vihman, Language and Linguistic Science, University of York, Heslington, York YO10 5DD, England (e-mail: marilyn.vihman@york.ac.uk).

Abstract

Phonological development is sometimes seen as a process of learning sounds, or forming phonological categories, and then combining sounds to build words, with the evidence taken largely from studies demonstrating 'perceptual narrowing' in infant speech perception over the first year of life. In contrast, studies of early word production have long provided evidence that holistic word learning may precede the formation of phonological categories. In that account, children begin by matching their existing vocal patterns to adult words, with knowledge of the phonological system emerging from the network of related word forms. Here I review evidence from production and then consider how the implicit and explicit learning mechanisms assumed by the complementary memory systems model might be understood as reconciling the two approaches.

In accounts of language development the phonological aspect is often overlooked. Yet its role is necessarily foundational: Knowledge of a certain minimum of words and phrases is an essential basis for learning grammar, and learning words depends on gaining knowledge of speech forms and of the links between those forms and their meanings, which must be deduced from their situations of use. The rapid advances in phonological development of the first 18-24 months are thus an important element in early word learning. Here I discuss the nature of those advances, the relation between development in the production and processing of speech sounds and whole-word units and the mechanisms that underpin human learning over the lifespan. The central questions that will guide the overview concern the first units and how they are learned: *Do infants begin by learning speech sounds and then combine them to recognize and produce words? Or do they begin by producing word-like vocalizations and retaining bits of the speech signal that match their production? Or do these processes occur in parallel?*

Before beginning on a review of developmental studies it will be useful to briefly characterize phonological structure, an essential part of the systematic knowledge that underlies fluent native language use in adult speakers. Consider two prominent contrasting views of adult phonology: The formalist view takes the segment or phoneme (or the bundle of distinctive features that make up the segment or phoneme) to be basic to linguistic structure (Chomsky & Halle, 1968; Halle, 1971; Blevins, 2004); the functionalist view sees units linked to meaning and communication (whole word forms) as basic to both phonological structure (e.g., Pierrehumbert, 2001, 2003; Port, 2007; Vihman & Croft, 2007) and speech production (Redford, 2015). Although phonemic oppositions undeniably play an important role in distinguishing linguistic units in any language, phonemes, the 'minimal units of distinctive sound function, forming a unitary inventory within a language and concatenated with one another in an additive way to form words' (Anderson, 1985, 292), are not the only key elements of phonology. The natural classes of segments that function in similar ways (but differently in different languages) – sometimes formally indicated through the use of distinctive features – express the paradigmatic axis of phonological structure above the level of individual segments, while prosodic structures, or phonotactic regularities, express the syntagmatic axis. For our purposes, then, phonology can be considered a *system of networks relating word forms along these two axes*.

At the same time, current phonological models often emphasize the dynamic nature of language in use in contrast to its idealized conceptualization. Thus, the exemplar models of usage-based phonology find that the same phonological sequences may have different ranges of phonetic variation in connected speech in apparently homophonous pairs of words with differing frequencies of use (cf., e.g., *four/for, can* (auxiliary)/*can* (main verb) and even, though less dramatically, *time/thyme, right/write,* etc.: Bybee, 2001; Pierrehumbert, 2001; Gahl, 2008). Furthermore, current work in sociophonetics has shown the importance, for explaining variability, of speakers' memory for word exemplars, which not only retain phonetic detail (rather than an abstract, minimally redundant sequence of phonemes) but also the socially relevant indices of the individual speaker that underlie accommodation and change (Foulkes & Docherty, 2006; Foulkes, 2010). This demonstrates the value, for robust processing, of richly specified representations as well as of multiple access routes, based on the network of connections built up with experience of individual items (Wedel, 2007; Menn, Schmidt & Nicholas, 2013).

Thus an understanding of how infants begin to gain knowledge of language need not start from the assumption that phonemes or individual speech sounds are the first elements of language structure to be learned. If there is a more direct 'phonetic' path to word-form knowledge out of which phonological structure may be taken to emerge in parallel with vocabulary growth, then Ferguson and Farwell (1975) were right to appeal to 'the primacy of lexical learning in phonological development' (112). I will consider evidence that supports such a direct route.

I begin by reviewing what is now generally understood to be the course of developmental advances in infant speech perception, from infants' remarkable early discriminative capacities to the 'perceptual narrowing' that marks their entry into the phonological structure of a particular ambient language. I will contrast two interpretations of this developmental course, one emphasizing the perceptual learning of phonological categories, the other, interaction with the parallel emergence of adult-like vocal production. I will then pursue an account of development from the perspective of production: Infants' well-practiced vocalizations (babbling) will be seen to support memory for similar sequences in the input speech stream ('pre-selection'); this leads to a small data-base of highly familiar word-forms or production routines, some of which then generalize to more abstract schemas or templates; these in turn attract and assimilate novel, more challenging word forms and thus facilitate further lexical development. Under this latter account knowledge of individual speech sounds is emergent from the process of systematization or networking of similar forms. Finally, I will consider the mechanisms that underlie the development of phonological knowledge in light of the complementary systems model current in the neuroscience of memory, which conceptualizes independent contributions from 'implicit' (distributional and procedural) and 'explicit' (declarative, attention-based) learning.

Infant speech perception and 'perceptual narrowing'

The study of infant speech perception was initiated with the finding that speech sound contrasts are discriminated categorically from the first months of life (Eimas, Siqueland, Jusczyk & Vigorito, 1971). Contrary to initial interpretations of this finding in terms of an innate human specialization for language, subsequent demonstration that chinchillas, macaques and dogs discriminate phonetic contrasts in the same categorical way suggested that special sensitivity to certain regions in the speech signal may be built into the auditory system that humans share with other mammals (Stevens, 1972, 1989); this led to the idea that categorical perception may reflect evolutionary auditory shaping of the phonology of human languages (Kuhl, 1986). Somewhat surprisingly, however, speech sound discrimination proved to be superior in infants, with their limited auditory experience, in comparison with the related function in adults, with their more narrowly circumscribed facility that discriminates just those contrasts that characterize their native language (Werker & Tees, 1983, 1984).

Infants' 'universal' discriminatory capacities in the first 6 months of life are well established: In the somewhat artificial conditions of repeated syllable presentation in a laboratory experiment infants readily discriminate consonantal contrasts and also vowels, whether they occur in the native language or not (see Vihman, 2014, ch. 3, for a review). The first perceptual 'advance', however, is an early *regression* in discrimination: Success, among groups of infants aged 6-8 months, in hearing differences between a variety of different speech sounds has been robustly shown to contrast with *failure* to discriminate phonological categories not distinguished in the ambient language by groups of infants aged 10-12 months. This shift in perceptual processing has been tested mainly in infants exposed to English (e.g., Werker & Tees, 1984; Best, 1994), but it applies equally well to infants exposed to other languages, such as Japanese (Kuhl et al., 2006) or Arabic (Segal, Hejli-Assi & Kishon-Rabin, 2016).

No fully satisfactory explanation for this early (and rapid) loss of a generalized capacity to detect segmental distinctions has been provided as yet. Instead, various plausible accounts have been offered, based on developmental shifts that occur around the same time. These include infants' emerging capacity for voluntary attention (due to maturational changes in inhibitory control: Tipper, 1992; Ruff & Rothbart, 1996), their dawning responses to meaning (Huttenlocher, 1974; Bates et al., 1979; Benedict, 1979; Bergelson & Swingley, 2012, 2013), their increasing skill in adult-like syllable production (Vihman, 1992; Davis & MacNeilage, 1995, 2000), their experience with speech sounds in the context of object labeling (Yeung & Werker, 2009) and their ongoing implicit learning of the statistical distribution of speech sounds experienced in the input without reference to meaning ('distributional learning': Maye, Werker & Gerken, 2002).

Speech sounds before words

Distributional learning is currently the most widely accepted source of the shift in perceptual discrimination. Maye et al. (2002) demonstrated 6- and 8-month-old infants' sensitivity to uni- vs. bimodal differences in distribution in a brief lab experiment: By editing and resynthesizing recorded tokens of English [da] and (unaspirated) [ta] they created eight CV-stimuli evenly spanning the acoustic continuum from voiced to voiceless unaspirated alveolar stops (cf. Pegg & Werker, 1997). They familiarized two groups of infants with these stimuli: One group heard more repeats of tokens in the middle of the range ('unimodal exposure'), the other group more repeats of tokens toward the extremes of the range ('bimodal exposure'). When tested with extreme tokens that had been presented to the same extent in both cases, only the infants provided with bimodal exposure discriminated the stimuli, at either age. This suggests that separate categories were formed only in that condition.

The potential relevance of this experiment for the issue of perceptual narrowing in speech sound discrimination is clear. Contrasting sounds can be expected to cluster separately in any language, with minimal overlap, while similar sounds that do not contrast are likely to be more diffusely distributed. According to this model, infants will naturally form phonological categories from denser clusters of sounds; contrasts falling outside of these categories will no longer hold their attention. This could explain why, for example, Japanese infants no longer discriminate English /r/ from /l/ by the end of the first year (Kuhl et al., 2006), Arabic-learning infants no longer discriminate Hebrew /p/ from /b/ (Segal et al., 2016), Urdu-learning infants no longer discriminate English /v/ from /w/ (Dar, 2016) and English-learning infants no longer discriminate the velar and uvular ejectives of the Interior Salishan (Native American) language, Nthlakapmx or Thompson (Werker & Tees, 1984) or the voiceless unaspirated and voiced labial stops of Zulu (usually described as voiced vs. implosive labial stops: Best & McRoberts, 2003).

Does this mean that infants begin by learning sounds and contrasts, and are only subsequently able to begin to register and represent word forms? The fading of discriminatory attention to infrequent or non-occurring category contrasts is accompanied by a sharpening of frequently experienced category boundaries, according to Maye et al. (2002) and Kuhl et al. (2006, 2008). Thus the growing strength of representation of individual phonological categories (contrasting segments or phonemes) has been taken to provide critical underpinnings for knowledge of word forms (Kuhl, 2004; Werker & Fennell, 2004).

On the other hand, unsupervised distributional learning is possible but demonstrably difficult for adults (Goudbeek, Swingley & Smits, 2009) and is insufficient in itself for inducing discrimination of some phonetic contrasts (Cristià, McGuire, Seidl & Francis, 2011). Furthermore, analysis of a good-sized corpus (700 single vowels produced by one mother to her 10-month-old) revealed a far greater extent of overlap in the distribution of distinct vowels than this model would predict (see Fig. 2, Swingley, 2009). Both Swingley and Cristià et al. conclude that learning based on acoustic cue distributions alone is unlikely to be sufficient to account for infant learning of the phonetic categories of their language (see also Werker, Yeung & Yoshida, 2012).

Recently, Yeung and Werker (2009) provided evidence that, when trained with consistent sound-object pairings, 9-month-olds – at a transitional age between 'universal' and 'native-language-only' discrimination of consonantal contrasts – are able to discriminate minimal contrasts in non-native consonants; without training, or with inconsistent exposure to the sound-object pairings, they fail to show discrimination. Thus the classical concept of 'acquired distinctiveness' (Bonardi, Graham, Hall & Mitchell, 2005) can be drawn on, alongside simple distributional learning, to account for the formation of phonological categories as children begin to attend more to the speech around them in relation to familiar objects and events.

Note that there are problems with this account as well, inasmuch as naturalistic speech to infants rarely involves a focus on objects or events labeled by minimal pairs. On the other hand, Heitner (2004) stresses the complementary and more plausible effect of growing lexical knowledge on phonological category formation: Although minimal pairs may be infrequent, within-category phonetic variants used for the same referent are abundantly available in any speech event, providing a highly serviceable means for children to form equivalence classes for potentially distinct sounds that they can accordingly learn *not* to discriminate.

Under any interpretation, however, the relationship of knowledge of phonological categories that might be gained through passive perceptual exposure to speech sound clustering in the input, with or without the support of meaningful reference, does little to account for infants' ability to produce sequences of speech sounds identifiable as word forms; no mechanism has been specified in the perception literature to indicate how the one source of learning might extend to the other – although vocal practice must presumably play a role. Idiscuss ways of understanding this relationship below.

If we assume that children are learning speech sounds within lexical contexts, the problem of phonological category formation purely from hearing speech becomes more tractable (Swingley, 2009, Fig. 3). However, the issue of how infants first learn to recognize word forms remains unresolved as well. There is ongoing debate as to whether they begin by picking up statistically frequent sequences, independent of any meaning function, and gain knowledge of the accentual system of the language based on that learning (Thiessen & Saffran, 2003, 2007), for example, or whether some aspect of prosodic (accentual) structure is primary instead (Johnson & Jusczyk, 2001; Johnson & Tyler, 2010).

An additional possibility, disfavored by most specialists in the area of infant word segmentation, is that the relatively small proportion of isolated word forms used in infant-directed speech – assessed at 9-10% of all words used if 'non-syntactic' words such as *uh-oh, wow, yum-yum* are disregarded (e.g., Brent & Siskind, 2001), a far higher proportion (40%) if such forms are included (Van de Weijer, 1998) – provides the infant with a 'wedge' into the speech stream (Swingley, 2009; Keren-Portnoy, Vihman & Lindop Fisher, 2015). Most pertinent for this discussion, isolated words offer a practical training ground for word-form learning. Evidence that such forms are an important source of early phonological knowledge is afforded by the fact that they number among the first five or six words identified for every one of the 48 children learning 10 languages listed in Menn and Vihman (2011).

Note that perceptual narrowing has also been found to occur, within the same time-frame, as part of category formation in a far broader range of cognitive domains, such as the discrimination of musical changes embodied in unfamiliar musical traditions or faces representing unfamiliar races (cf., e.g., Scott, Pascalis & Nelson, 2007; Lewkowicz & Ghazanfar, 2009; Maurer & Werker, 2013) or species (Pascalis, DeHaan & Nelson, 2002). Within these quite different domains, maturational as well as experiential changes in attentional capacities must be relevant alongside any distributional factors.

We can assume that such broad attentional shifts are linked to developmental changes in what is meaningful for the infant, socially or affectively as well as referentially or semantically (i.e., in relation to word meanings). For example, infants fixate on faces for the first few months of life; their growing knowledge of faces, combined with the powerful emotional experiences associated with them, is a critical part of the process of widening social engagement, a foundational aspect of being human (Boysson-Bardies et al., 1993).

Similarly, infants' advances in experience of 'action' or purposeful movement, which support their growing sensorimotor knowledge of the physical world, also support conceptual advances (Thelen & Smith, 1994). All of this must be involved as well in the apparent category formation that results in perceptual narrowing, inasmuch as the linked cycles of (self-) action and perception have been shown to underlie so much of cognitive and social as well as motoric development (see Campos et al., 2000). Thus the emergence of adult-like vocal production in the middle of the first year of life, in the form of the first CV syllables ([bababa, dadada, ŋaŋaŋa]), could be expected to affect infant speech processing as well, focusing infant attention on selected (matching or sufficiently similar) portions of the input speech stream and thus potentially playing a role in the fading of the early 'universal' capacity to discriminate phonetic differences.

Whole words before speech sounds

An alternative theoretical approach to phonological development is to assume that children do not learn speech sounds directly at all. Instead, they learn whole word patterns, with knowledge of those speech sounds frequently experienced in familiar words later emerging out of the representational network of known words of similar length, accentual pattern and/or onsets, rhymes and codas. This assumption derives primarily from production studies, which provide ample evidence that the first words are typically learned as whole items or sound-patterns (Vihman & Keren-Portnoy, 2013). However, Werker and Curtin (2005), whose PRIMIR model of phonological development draws primarily on perception studies, nevertheless similarly propose that 'once the infant has established a sufficient number and density of meaningful words, generalization of commonalities occurs, leading to the emergence of the Phoneme plane' (214).

A related conceptualization of a network of connections emerging from individually known lexical items underlies exemplar models of phonology (Bybee, 2001, 2010; Pierrehumbert, 2001; Wedel, 2007). Similarly, Edwards, Munson and Beckman (2011) see phonology as emerging from 'generalizations over the parametric phonetics and generalizations over the lexicon' (37; see also Beckman & Edwards, 2000a, b; Munson, Edwards & Beckman, 2012).

The conceptualizations of phonological development as beginning with sounds or with whole words seem to clash, yet there is good reason to believe that each of these accounts is at least partially correct (see also Swingley, 2009). How might the evidence from perception studies that supports the early distributional learning of speech sounds be reconciled with the evidence from production studies for whole-word learning? I will review the evidence from production studies of the first 18 months. I will then consider what learning mechanisms might be able to account more satisfactorily for the evidence from both perception and production studies, based on current work in neuroscience.

Word production, I: Item learning and 'pre-selection'

Infants' early capacities for perceptual discrimination, shared with other mammalian species, contrast sharply with the uniquely human ability to produce the core syllables basic to the phonology of the world's languages. That ability is absent at birth but develops rapidly, typically appearing in identifiably adult-like vocal production by 6-8 months (Oller, 2000), with little apparent variation across ambient languages. This key production milestone is followed within a few months by an emergent capacity to represent, recall and produce familiar word forms. More specifically, the first words are identified about four months after the onset of 'canonical babbling' or CV syllables, at the earliest (based on parental report for 18 infants: Fagan, 2009).

Untrained word-form recognition (without the support of visual images), which reflects long-term representation of words heard frequently in everyday life (DePaolis, Vihman & Keren-Portnoy, 2014), may precede full word comprehension (Hallé & Boysson-Bardies, 1994; Swingley, 2009). There is no behavioral evidence of such recognition at 9 months, while at 10 months the experimental effects are variable and related to infant production experience: Infants who are consistently and stably producing one or more consonants in repeated recordings perform at the extreme end of a scale of 'preference ratios' (proportion of looking time to common or 'familiar' words out of total looking time) – either showing a familiarity or a novelty response; those not yet producing consonants consistently perform at chance, showing no significant preference for either set of words (DePaolis, Keren-Portnoy & Vihman, 2016). By 11 months untrained word-form recognition is robust (Vihman, Nakai, DePaolis & Hallé, 2004; Vihman, Thierry, Lum, Keren-Portnoy & Martin, 2007).

Experimental studies reveal differences by ambient language in the aspects of word forms that hold infant attention, such as onset consonant of the accented syllable in English and French (Vihman et al., 2004) or medial geminates (phonologically contrastive long consonants) in Italian (Vihman & Majorano, in press). This testifies to the shaping effect on attention to speech of exposure to input over this period, when babbling becomes an increasingly dominant feature of social interaction (cf. Goldstein, King & West, 2003).

Both in infancy and beyond, studies have shown that word production and use provide a more stable, more reliable, better-established representation than word recognition or comprehension alone (MacLeod et al., 2010; Vihman, DePaolis & Keren-Portnoy, 2014; Icht & Mama, 2015; Zamuner, Morin-Lessard, Strahm & Page, in press). There are many reasons why this should be true, including the greater effort involved in production, which accordingly supports more robust memory or representation (Elbers & Wijnen, 1992) and the support that a match to a well-practiced production routine affords to the challenge of retaining novel word forms; the matching process, which becomes increasingly accessible as the lexicon grows, constructs or shapes phonological memory (Keren-Portnoy et al., 2010).

Furthermore, expressive vocabulary constitutes a strong predictor of lexical advance: What is known already affects the way the brain processes what is new. This has been shown indirectly in studies of processing speed (e.g., Fernald & Marchman, 2012) and eye-tracking (Horváth, Myers, Foster & Plunkett, 2015) as well as in direct measurement of brain function (Torkildsen et al., 2008, 2009). *How are these various advances interrelated? And how does the emergent function of speech-like production relate to the ability to process speech sounds?*

The idea that emergent control over vocal production might affect infants' processing of speech was initially proposed as a way to account for the fact that infants' earliest words are surprisingly accurate (as first noted by Ferguson & Farwell, 1975). That is, the first words may show some omission or substitution of consonants but, generally speaking, they constitute simple matches to comparably simple one- or two-syllable target words, as illustrated in Table 1 (see Appendix I, Menn & Vihman, 2011).

[Insert Table 1 about here.]

The phonetic repertoire seen in the first words is the same as that which characterizes babble, which is widely accepted to be unconscious practice for word production (Vihman et al., 1985). As seen in Table 1, the consonants are largely restricted to stops and nasals, glottals and glides and the forms rarely include more than a single supraglottal consonant type. Aside from the voicing contrast in Annalena's [data] for *das da*, the only exceptions are Alice's [m̩ːɑnːə] for *mommy* and Kaia's [ki̥ːtɔ̥] for *kiisu*. In the remaining words in Table 1 with more than a single 'true' (supraglottal) consonant in the adult form we find consonant harmony, or full consonant agreement across the word, either in the child form alone or in both child and adult forms.

Menn and Vihman (2011) comment on the relatively unsystematic nature of the first word forms seen in almost half of the children whose data they present: Those children 'seem to have acquired a word-length complex of gestures as an unanalyzed whole' (271). In other words, these children give evidence of learning words – picking up on the occurrence of something similar to their existing vocal forms in the input, with no need for analysis – before learning speech sounds. In general, infants' first words are similar to babble and build on that vocal practice, which strongly predicts vocabulary growth (McGillion et al., in press); the relative accuracy of these words suggests 'pre-selection' (Ferguson & Farwell, 1975) and the lack of systematicity suggests that this is not yet 'phonology' (Vihman & Velleman, 2000).

The relationship of emergent vocal production skills to speech processing: The articulatory filter hypothesis

The presumed phenomenon of pre-selection calls for explanation. To account for it Vihman (1993, 1996) proposed that, once children have begun to produce adult-like syllables on a regular basis, they may experience as particularly salient those frequently heard input forms that resemble whatever is most frequent in their own vocal output. In other words, building on their ongoing experience with babbling, children begin learning words by retaining in memory whole adult forms that resemble their own frequent vocal patterns – although the resemblance will not be an exact acoustic match, given the differences in adult and child vocal tract sizes and proportions.

Evidence for the plausibility of such an auditory matching process in the pre-lexical period comes from Goldstein and Schwade's (2008) study of 9-month-old infants: In their 'contingent maternal response' groups, infants responded with vowels to maternal 'vowel-only' utterances and with an increase over baseline in syllabic (CV) vocalizations to maternal syllabic utterances; however, they were found not to imitate the mothers but only to produce forms that 'matched' the global characterization, V to V sequences and CV to CV sequences. This experimental finding illustrates, in a tightly controlled condition of exposure to isolated maternal speech forms, an early phase of infant holistic or 'whole-form' matching, in which adult speech forms shape vocal responses at an age when canonical babbling is normally well established but first words are not yet typically identified. (Messum & Howard, 2015, provide an alternative interpretation of such interactions, with adult mirroring rather than infant responses to a match serving to teach the child equivalences between their production and adult speech forms.)

Vihman (1993, 1996) referred to the concept of selective infant attention to broadly matching adult speech forms as the 'articulatory filter', to express the idea that the child unconsciously filters what she hears in the input through her own production experience. The proposal is justified in part on the grounds that the child's own vocalizations, themselves guided or primed by often-heard words or patterns in the speech stream (Boysson-Bardies & Vihman, 1991), will have a double effect on the child, being experienced as both an auditory and a proprioceptive stimulus (Vihman et al., 2014). This should strengthen the child's representation of speech forms that resemble her typical production (i.e., forms that constitute a rough match of input and 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. (Vihman, 1996, 142)

This hypothesis remained purely speculative until DePaolis, Vihman and Nakai (2013) ran an experimental study of 53 children acquiring either English or Welsh in North Wales. DePaolis et al. recorded infant vocalizations in the home four times over a two-month period, beginning at 10.5 months. Two weeks after the last session, at about 12.5 months, they tested the infants in the lab, presenting them with two lists of nonwords, each making repeated but varied use of one of two supraglottal consonants that are equally frequent in the input but that are expected to differ in extent of child use in production at this age ([t] vs. [s] for English, [b] vs. [g] for Welsh).

The results were consistent with the hypothesis that production affects speech processing, but in an unanticipated way: The English children with the highest production of [t/d] in the final session showed greater interest in /s/, the speech sound they were producing only rarely, if at all, than in /t/, the speech sound they were most familiar with through production; only children with lower [t/d] production in that session showed greater interest in the stop than in the fricative. The Welsh children, whose production was not well differentiated for the two speech sounds tested, showed roughly the same level of interest in both sets of stimuli.

This experiment demonstrated for the first time an effect of infant vocal production on speech processing, however paradoxical. Two subsequent experiments made use of the individual differences consistently seen in infant vocal production to test the articulatory filter idea more directly. In order to more specifically test whether infants match their own patterns to input speech these studies adopted from McCune and Vihman (2001) a measure of consistency (or identifiability) and stability of vocal production, the 'vocal motor scheme' (VMS), which picks out recurrent and stable speech-sound use.

DePaolis, Vihman and Keren-Portnoy (2011)[1] recorded infants in their homes, beginning at 9-11 months, and transcribed the sessions as quickly as possible to permit timely testing, as soon as evidence of consistent, stable use of a single speech sound (VMS) emerged. Eighteen infants met the criteria for such use in the home recordings and were tested on short passages featuring nonwords with either a VMS the child was using ('own VMS'), a different possible VMS that the child was not using ('other VMS'), or a labiodental fricative, to control for the effect of a speech sound none of the children were likely to be using with any frequency ('non-VMS').

The experiment was subsequently replicated with 26 Italian children, first seen at around 6 months and then recorded in the home longitudinally from the onset of canonical babbling (between 7 and 11 months) until at least one VMS was identified (Majorano, Vihman & DePaolis, 2014). In the Italian study infants were tested with word lists (as in DePaolis et al., 2013), similarly contrasting 'own-', 'other-' and

[1] Although the 2011 study was published earlier, the 2013 study had actually been run some years previously.

‘non-VMS’. As the findings of the British and Italian studies are similar I report them together here.

The results were consistent with the English/Welsh study, in that the infants fell into two groups, depending on their level of VMS knowledge. Those with more than one VMS in repertoire were significantly more interested in ‘other-’ than in ‘own-VMS’ (no group differences were found in relation to ‘non-VMS’; that condition is not further discussed here), while in the larger Italian study infants with a single VMS were significantly more interested in their own VMS (in the British study the same effect was only a trend): See Figure 1. Note that in both groups the Italian infants looked longer at both sets of stimuli than did the British infants: This presumably reflects the difference in presentation, with the individual VMS being more readily accessed in word lists (Italian) than in the passages from which the VMS-rich words had to be segmented (British). In addition, the Italian study found in a separate experiment that in the pre-VMS period (at 6 mos.) there was no difference in attention to the different stimuli, which were distinguished only by their subsequent VMS status for the child and which were thus not expected to affect processing at an earlier developmental point.

[Insert Figure 1 about here.]

The findings of the three studies are in good accord. When a child first begins producing one consonant stably and consistently, as established by VMS identification, he or she is particularly attentive to that speech sound in input word forms (as shown in Majorano et al., 2014). When the child has advanced to production at VMS level of more than a single consonant the known (VMS) consonants no longer hold his or her attention; instead, the child seems to discover a world of varied stimuli and to begin to attend more to what is novel or unfamiliar (for further discussion see Vihman et al., 2014; DePaolis et al., 2016). This series of studies solidly establishes an effect of the child’s own level of production of speech sounds on the way she processes or represents those sounds.

Thus the experimental evidence supports the pathway from vocal practice to first words that I have proposed based on observational findings (Vihman, 1993, 1996). First, babbling lends salience to aspects of the input. In exemplar theory terms, the similarity between heard word-form and existing child vocal pattern creates an ‘echo’ or resonance (Goldinger, 1996, 1998). As a result, secondly, frequently heard word forms come to be represented more robustly in the child’s mind than forms for which the child lacks a possible vocal match. Thirdly, production-based salience in the speech stream facilitates formation of a form-meaning link in relevant and frequently repeated contexts. This can in turn result in early identifiable word production, under priming from a familiar situation of use. This account, which sees in typical early word production individual, unrelated instances of ‘item learning’ based on infant sensitivity to rough matches between simple target structures and their own motoric routines, accounts for several well-established characteristics of children’s first words: their accuracy, their similarity to the particular child’s babbling repertoire and their typical lack of systematic phonological relationships.

Word production, II: From holistic matches to reorganization and systematicity

‘Holistic’ representation means memory for the word-form as a whole, with retention of salient syllables or segments, particularly those that are within the child’s output repertoire, but not necessarily for the exact segmental sequence or for every aspect of the form; furthermore, ‘holistic’ representation implies that elements in one part of a word may affect memory for other parts. Such holistic representation is suggested by the findings of the untrained word-form recognition studies mentioned above. At 11 months infant long-term memory for word forms is robust for lists of accurate word forms but is blocked when certain aspects, such as onset to the accented syllable, are changed or ‘mispronounced’. On the other hand, change to other aspects of such words fails to block word-form recognition, implying that those aspects were poorly represented. Thus, final consonants of monosyllables in Dutch (Swingley, 2005) or onset consonants in unaccented syllables in French (Hallé & Boysson-Bardies, 1996) or English (Vihman et al., 2004) appear to be weakly represented in the pre-linguistic period. Similarly, Vihman & Majorano (in press) demonstrate the perceptual neglect of word-initial consonants in Italian words with medial geminates, though not in those with medial singletons – suggesting that the medial geminate may draw attention away from the accented word-initial syllable.

All of these experimental findings reveal that some elements of early word forms are less well represented than others and suggest that the form of the word as a whole affects infant processing. In fact, word production studies provide ample evidence of a difference between strongly represented sounds, such as the word-initial consonants of trochaic (strong-weak) words in English, which are rarely omitted, and weakly represented sounds, such as the word-initial consonants of iambic (weak-strong) words in French (Vihman & Kunnari, 2006) or Hebrew (Keren-Portnoy & Segal, in press) or the onsets of trochaic words with geminates in Estonian, Finnish and Hindi as well as Italian (Vihman, 2016; Vihman & Croft, 2007; Vihman & Majorano, in press), all of which are commonly omitted in child word forms in the single-word period.

First word use leads to a small expressive lexicon. Once a few different words are being used with some regularity, children are typically found to generalize, or to begin to overuse, one or more of their production patterns, with two effects on their word forms taken as a whole: (i) they become more similar to one another and (ii) they become less accurate. In other words, I see regression in match to the adult model as a concomitant of the advance in systematicity evidenced by the increased similarity of the child’s forms, which begin to fall into a small number of often-used prosodic structures (or overall word structure in terms of length in syllables and of consonant and vowel [C-V] sequences). Such favored child word patterns are termed phonological templates, idiosyncratic child patterns found to apply both to ‘selected’ words, which exemplify the pattern, and ‘adapted’ words, more challenging adult word forms that are assimilated to it (see Table 2).

[Insert Table 2 about here.]

Table 2 illustrates the templates of the four children whose first words we saw in Table 1. For three of these children something of the later template can, in hindsight, be identified in their first words: Annalena is extending a preference for reduplicated forms to more complex targets (Vihman & Croft, 2007), Alice is building on an affinity for producing words with palatal consonants and final [i] (Vihman et al.,

1994) and Laurent has systematized his use of [l], now producing words in which it serves as the onset to the accented syllable, regardless of the actual structure of the word or phrase he is targeting (Vihman, 1993). In Kaia's case, however, only one of the first words includes a medial geminate, the basis for her template at 16 months (Vihman, 2016).

To illustrate just how templates function, consider Annalena's forms for her name and for *Zahnbürste*, both produced as [nana], even though strictly speaking /nana/ occurs as a sequence in neither word. 'Finding' such a sequence in target forms like these must be due in part to the existence in the child's mind of a schema or template of the shape 'repeated syllable' or <$\square_1$ $\square_1$>. The order reversal (metathesis) that we see in *Zahnbürste* is not unusual in words adapted to fit a child's template in this period of lexical development.

Use of templates reflects a child's generalization of production patterns as their word learning advances. This can be conceptualized in at least two different ways: (i) as a purely procedural or motoric extension of existing production routines; (ii) as 'secondary' distributional learning (Vihman, 2014), based on each child's individual database of early words. In either case the template is necessarily shaped by the ambient language target forms as well as by the child's individual production patterns. The choice of theoretical conceptualization is independent of the data themselves, which are robust: Evidence of template formation, at varying levels of lexical development and for varying periods of use, is available for a range of different languages – all those, in fact, for which individual cases of phonological development have received close linguistic analysis, although not all children provide evidence for such patterns (see the seven languages represented in Vihman & Keren-Portnoy, 2013, an overview of 13 languages in Vihman & Wauquier, in press, and an analysis of typological differences between templates in 44 children learning English, French, Italian or Finnish in Vihman, in press).

Regardless of the mechanism, it is clear that these favoured routines or templates facilitate production – including articulation, planning and memory, or access to an emergent, still unstable representation. McGregor and Johnson (1997) put it succinctly:

> Template application allows the child to fit a production to a well-practiced routine, thereby reducing the demand on resources. Templates may aid the memory for the sound system as well as the planning and execution of motoric gestures. (1220)

Learning mechanisms: The complementary systems model

How does the initial attunement to the native language described in the first part of this paper, the decline in attention to non-native contrasts based on passive exposure to speech, relate to the attention-based item learning that I have discussed and illustrated with children's first words? Infant knowledge of speech sounds based on distributional learning cannot account for the production of identifiable word forms. In contrast, experience of vocal production and word use *can* give rise to implicit knowledge, for the purposes of perceptual processing as well as for production, of the phonological categories of the ambient language. Based on the idea of 'dissociated memory systems' (Schacter & Moskovitch, 1984) and its subsequent development into the complementary learning or memory systems model (McClelland, McNaughton & O'Reilly, 1995; O'Reilly & Norman, 2002; Lindsay & Gaskell, 2010), knowledge of both words and sounds can be understood as the byproduct of the integration, in active word learning and use, of implicit and explicit learning mechanisms. (See Ellis, 2005, for a similar account of L2 learning.)

Very few experimental studies have directly addressed memory functions in relation to infant word learning (but see now Friedrich, Wilhelm, Born & Friederici, 2015; Horváth et al., 2015). However, studies of word learning in adults and older children (e.g., Dumay & Gaskell, 2007, 2012; Backhaus et al., 2008; Henderson, Weighall, Brown & Gaskell, 2013; Brown & Gaskell, 2014; Gaskell et al., 2014; Takashima et al., 2014; Henderson, Devine, Weighall & Gaskell, 2015) demonstrate the applicability to this domain of the principles of the complementary systems model.

The memory system must be plastic enough to allow new learning, yet new learning must not be allowed to overwrite existing knowledge (the 'stability – plasticity dilemma'). The proposed solution is learning supported by two independent brain systems (Kumaran & McClelland, 2012; McClelland, 2013): (i) The *neocortex* gives rise to 'incidental', implicit (including distributional or statistical) or procedural learning, with no need for focused attention; (ii) the *hippocampus and the prefrontal lobes* together support learning *with* attention (Wilhelm, Prehn-Kristensen & Born, 2012). Note, however, that neocortical activity is always present, whether focal attention is also engaged or not; this is one of the many difficulties involved in assessing the independent contribution of each of the two systems to subsequent access, (implicit) recognition and (explicit) recall (Jacoby, 1991).

Implicit (distributional or statistical and procedural) learning

The sensorimotor areas of the neocortex *learn slowly from repeated experiences*, gradually gaining automaticity in motor skills (procedural learning, such as balancing on a bicycle or producing a particular vocal form), tallying statistical co-occurrences and, crucially, *categorizing the new in terms of what is already known* (including 'secondary distributional learning', discussed below); only minimal attention, if any, is required for this incidental experiential learning. Implicit learning of any kind supports unconscious, involuntary recognition and a 'feeling of familiarity' when previously experienced items or events – or items or events that closely resemble what was previously experienced – are encountered anew (Jacoby, 1991). Access to such implicit or procedural memories is possible *only* with close contextual matching, however; it is not available to consciousness and cannot be called up at will.

Explicit or declarative learning

One function of the prefrontal lobes is to focus attention on aspects of experience and inhibit attention when it is no longer required, permitting the kind of flexible selection of points of focus that begins to appear in infants only from the second half of the first year (Ruff & Rothbart, 1996). Together with the prefrontal lobes – which strictly channel experience, permitting only a single focus – the hippocampus serves to bind the experienced event together with all of its unique spatiotemporal features; in adults it is the key mechanism for retaining in memory the conjunction of separate (multimodal) aspects of experience. This notably includes the most essential characteristic of human language, the (typically arbitrary) link between a speech form and its situational context or meaning in a particular instance of use (an episode).

These hippocampal snapshots of episodes experienced with attention underlie spontaneous (conscious, voluntary) recall; this is the kind of memory generally required for item learning.

At least some parts of the hippocampus are known to be immature in the first few years of life, based on studies of visual processing in monkeys and to some extent also human infants. Some memory functions thought to be hippocampus-dependent are in evidence from the first months, however; these include 'recognition memory', based on visual paired comparison (VPC), or familiarization with a visual stimulus, followed by testing with both the familiar and a comparable unfamiliar stimulus; this elicits a novelty response from the first months of life (Fagan, 1977).

Richmond and Nelson (2007) sketch out the likely developmental profile of hippocampal involvement in declarative memory advances in infants, with specific reference to encoding, retention and retrieval. Based on infant studies of deferred imitation and the VPC in relation to what is known from animal studies and studies of amnesia in adults, Richmond and Nelson suggest that maturation of the hippocampus alone is unlikely to be responsible for these advances. They find myelination, or the insulation of axons in the central nervous system, to be an important factor in improvements in processing speed in encoding, while the last part of the hippocampal complex to mature, the dentate gyrus, is particularly important for retention and retrieval. This critical brain area reaches its peak cell numbers and synaptic density at 16-20 months, with pruning to adult-like levels by 3 or 4 years (Huber & Born, 2014). Most relevant here is the finding that infant memories are initially highly specific but become less constrained with age: See, for example, Robinson and Pascalis (2004), who showed that VPC *with contextual change between familiarization and test* is possible at 18 but not at 12 months.

The 'representation of arbitrary…relations among the constitutional elements of an event' (Jabès & Nelson, 2015, 296) has been shown to be present in the visual domain by 9 months of age (Richmond & Nelson, 2009), with familiarity responses to repeated images of faces presented against the same or a changed background; long-term memory was not tested. For comparison, note that infants' memory for words known from the home, not trained in the lab, is reliably seen experimentally at 11 (but not 9) months (Vihman et al., 2004), despite the fact that everything about the test situation is unfamiliar – the voice presenting the stimuli, the darkened test booth, the disembodied speech – and no contextual information is available to prime recognition. Comparable long-term representation of word forms embedded in sentences emerges only about a month later (DePaolis et al., 2014).

However, advances in infant knowledge gained from actions in the world play a key role in memory function as well (Richmond & Nelson, 2007). Herbert, Gross & Hayne (2007) provide direct evidence of an effect on memory retrieval of experience with action, and similar effects of previous experience or knowledge on new learning can be inferred from several other studies. For example, DePaolis et al. (2016) relate vocal production to word-form recognition at 10 months, Horváth et al. (2015) show, at 16 months, a correlation between reported expressive vocabulary size and novel word learning and Fernald, Swingley and Pinto (2001) demonstrate that expressive vocabulary size, not age, is the best predictor of processing speed at 18 and 21 months.

Jabès and Nelson interpret the findings from studies of relational memory for visual events, which shows steady improvement over the first two years, as suggesting that

> infants might learn the relation between items and their context, but …this relational representation is unitary at first…such that retrieval is disrupted if components of the learned event are changed between the learning and test phases. In other words, relational memory is at first extremely specific to the context in which learning occurs and gradually becomes more "flexible"… allowing the generalization of learning to other conditions…This flexibility is a fundamental component of relational memory, which is thought to depend on the integrity of the hippocampal formation. (297)

Evidence from word production, which shows flexibility (generalization to new contexts) early in the second year but seldom before, fits well with these ideas about advances in declarative memory, although no direct studies of the neurological structures supporting this functional change have yet been carried out. A shift from 'context-limited' to 'context-flexible' word use has long been reported for children in the second year of life, with more flexible use being apparent from about 14 months on (Bates et al., 1979; Vihman & McCune, 1994; McCune, 1992, 2008; McCune & Vihman, 2001). Specifically, the first words are typically produced in situations that prime them, such as daily routines. Evidence of the extension of word forms to novel situations of use (e.g., saying *monkey* in response to an unfamiliar image of a monkey, or *uh-oh* on encountering an unanticipated, novel misadventure or change) signals the onset of *referential word use,* in which words begin to be used as symbols, not as parts of a contextually bound experience (saying *monkey* in response to the child's own monkey puppet or *uh-oh* upon routinely dropping an object).

To return to the complementary memory systems, how are they thought to work together? The hippocampus supports rich but sparsely distributed neural codes, which are resistant to interference between similar experiences, thus supporting very specific, concrete memories for episodes; the neocortex, in contrast, abstracts the structure underlying related experiences through its use of overlapping codes (Kumaran & McClelland, 2012). A key function of sleep appears to be the deeper processing of experiences, with active intercommunication between the hippocampus and the neocortex (e.g., Walker & Stickgold, 2004), in children as well as adults (Backhaus et al., 2008). In sleep, through neural reactivation of elements of experience, attention-based memory traces are restructured and consolidated (Lindsay & Gaskell, 2010), resulting in categorization into networks of sound and meaning. The process of selective strengthening of associations involved in this restructuring may be a key factor in the 'discovery of a shared structure' in representations (Drosopoulos, Schulze, Fischer & Born, 2007). In other words, the process of integrating new experiences with what is already known may constitute the critical basis for the generalization of knowledge that yields phonological categories and systems.

The term 'secondary distributional learning' (Vihman, 2014) can be used to characterize the generalization or abstraction of schemas or templates from word forms familiar from production, resulting in the child attempting more challenging word forms, but with a loss of accuracy. Thus the occurrence of templates, typically first observed in the earliest period of referential word use, can be seen as expressing the child's growing ability to generalize, through an implicit comparison of forms or exemplars, from what is known – individual internal representations of their own often-produced vocal patterns or word forms – to what is unknown. This seems to be a plausible effect of the integration of implicit and explicit memory functions.

More research is needed to explore these parallels and the possible role of differing aspects of brain function in lexical and phonological development, but it is now beginning to be possible to relate our understanding of cognitive development, once set out in terms of monolithic shifts from one 'stage' to another (Piaget, 1951, 1952, 1954; see the critique in Thelen & Smith, 1994, Ch. 2), not only to the dynamic advances and changes, based on the interaction of action and perception, that occur in different ways in different children, but also to their possible underpinnings in brain development.

Integrating the findings

After this excursus on the learning or memory mechanisms available for constructing knowledge I return to our opening questions: *Do infants begin by learning speech sounds and then combine them to recognize and produce words? Or do they begin by producing word-like vocalizations and retaining bits of the speech signal that match their production? Or do these processes occur in parallel?*

We can see templates as mediating between input- and output-based learning and as constituting a first step in the development of knowledge of both paradigmatic and syntagmatic phonological structure: First, over the first several months of life, the child becomes familiar with input speech, which comes to include his or her own adult-like vocalizations or output forms. This familiarization process itself can be understood as involving two processes occurring in parallel: (i) Statistical or distributional learning, which operates as early as 6-9 months (Saffran, Aslin & Newport, 1996; Thiessen & Saffran, 2007), provides growing familiarity with the overall ambient language structure; (ii) at about the same age the child first begins to recognize words, as regards either form or meaning (Tincoff & Jusczyk, 1999; Vihman et al., 2007; Bergelson & Swingley, 2012). This emergent knowledge can be understood as corresponding to clouds of exemplars of similar forms for frequently heard words or short phrases, with indexical aspects of both form and meaning also included, such as the particular speaker's voice and elements of the associated event or context (Jusczyk, 1997; Houston & Jusczyk, 2000, 2003; Pierrehumbert, 2001, 2003).

Note that although babble is produced in strings of varying lengths, target-based word production is commonly limited to one or two syllables, regardless of ambient language structure (Vihman & Wauquier, in press). This limitation is most likely due to infants' untutored phonological memory, which will come to retain longer and more complex input elements as a concomitant of growth in lexical experience and use (Keren-Portnoy et al., 2010). Thus, while a general sense of the prosodic, phonotactic and coarticulatory regularities of the ambient language is gained by the end of the first year, as shown in segmentation studies (see Vihman, 2014, Ch. 5), lasting traces of individual lexical forms (exemplars) can be expected to accumulate more slowly and to be limited by strong representational constraints on word length and complexity.

Once the child begins to produce his or her own word forms with specific targets, the same implicit mechanisms (distributional learning of sequences and patterns, self-organization of exemplars) can be assumed to operate in combination on the new database formed from the child's own words. The increasing numbers of representations of forms the child is producing, albeit with a good deal of variability, will at some point generally become robust enough – in combination with the relatively slow pace of advances in neuromotor control and speech-planning (cf., e.g., Payne, Post, Astruc, Prieto & Vanrell, 2012) – to give rise to one or more templates, although individual differences in child 'tolerance for variability' (Kamhi, Catts & Davis, 1984), or willingness to attempt challenging targets, will determine the extent of adaptation to templatic patterns (Vihman, 2016). As the child shifts from a primarily outward- to a primarily inward-oriented model for production we see the regression in accuracy described above along with an increase in the numbers of different word types produced.

This account points to the lexicon as the source of longer-term, robust phonological knowledge of individual segments (the Phoneme plane of the PRIMIR model: Werker & Curtin, 2005). The representations of production units, or units of form that have a link with meaning, can be expected to self-organize into networks based on similarity. Evidence of infant reliance on such networks can be seen in child lexical selection errors or 'mini-malapropisms', which tend to be based more often on holistic word-form similarities such as length in syllables and accentual pattern than on agreement in the initial sound, the most common basis for adult errors of this kind (Aitchison, 1972; Fay & Cutler, 1977; Vihman, 1981).

The relations between sub-units in different words, whether word-initial consonants, rhymes, accentual patterns or other repeatedly represented elements, are subsequently analysed implicitly (for accounts of longitudinal change that suggest such a process of reorganization, see Priestly, 1977; Macken, 1979; Vihman & Vihman, 2011). As Edwards, Munson and Beckman (2011) put it, 'phonemes do not exist in nature, to be "discovered" by children. Rather, they emerge gradually as children make increasingly robust abstractions over the words that they learn' (38). In short, self-organization and implicit analysis mean systematization and integration into networks of phonological similarity. Those networks provide multiple access paths to shared 'positional variants' (Pierrehumbert, 2003) or phonemes, strengthening the representation of speech sounds with every instance of language use, whether receptive or expressive.

The characteristic profile for growth in word comprehension is a slow start (by 6-9 months at the earliest: Bergelson & Swingley, 2012) followed by a rapidly rising curve, with a first inflection being observed only at about 14-18 months (Oviatt, 1980; Bergelson & Swingley, 2012, 2013), despite the fact that word-form recognition is reliably seen, cross-linguistically, by 11 months, as indicated above. The gap between initial word comprehension and word-form recognition and the more rapid, steadier advances in lexical learning that follow presumably reflects the benefit, for novel word learning, of a growing reference sample of familiar forms to which the novel items can be connected. Phonological memory, which develops through the emergent use of word forms in production (Keren-Portnoy et al., 2010), can be taken to be a key element here. In addition, the memory studies offer an account of how novel experiences are restructured through assimilation to existing patterns, which clarifies the importance of existing knowledge for consolidating new advances. This would support the idea that the onset of word production plays a pivotal role in integrating (i) emergent infant familiarity with the phonological categories of the ambient language with (ii) the infant's growing receptive lexicon of form-meaning pairs.

Flexible word use – that is, the child's spontaneous use, outside of priming contexts, of word types generalized from particular instances – is also typically observed from about 14 months (see Vihman, 2014, Fig. 6.1). Finally, the phonological template use that I have illustrated here, involving generalization or schema formation rooted in a learned database, is most often observed from about the same age, although template emergence, as indicated above, is tied to lexical practice and growth, but not in any predictable or mechanistic way; individual differences prevail here, not universal rules or stages.

In short, as infants begin to gain knowledge of a small number of often heard words (the first attention-based item-learning) and to register (implicitly) differences in the distribution of phonological categories in input speech, they are laying the foundation for first word production, which additionally depends on babbling practice. Production of some 50 to 100 different word types, in turn, prepares the ground for more rapid learning of new words, supported by an emergent capacity to generalize both form patterns (as initially seen in templates) and meanings (as seen in referential or symbolic word use). Given this conceptualization of knowledge and learning there is no real clash of sounds-before-words vs. words-before-sounds: The learning of sounds and words necessarily proceeds in parallel.

References

Aitchison, J. (1972). Mini-malapropisms. *British Journal of Communication Disorders, 7,* 38-43.

Anderson, S. (1985). *Phonology in the Twentieth Century: Theories of rules and theories of representations.* Chicago: University of Chicago.

Backhaus, J., Hoeckesfeld, R., Born, J., Hohagen, F. & Junghanns, K. (2008). Immediate as well as delayed post learning sleep but not wakefulness enhances declarative memory consolidation in children. *Neurobiology of Learning and Memory, 89,* 7-80.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Volterra, V. (1979). *The Emergence of Symbols: Cognition and communication in infancy.* New York: Academic Press.

Beckman, M. E., & Edwards, J. (2000a). Lexical frequency effects on young children's imitative productions. In M. B. Broe & J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press.

Beckman, M. E., & Edwards, J. (2000b). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development, 71*, 240-249.

Benedict, H. (1979). Early lexical development: Comprehension and Production. *Journal of Child Language*, *6*, 183-200

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 Sciences of the USA,* 109, 3253-3258.
Bergelson, E. & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition, 127,* 391-397.
Best, C. T. (1994). The emergence of language-specific phonemic influences in infant speech percpetion. In J. C. Goodman & H. C. Nusbaum (eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge, MA: MIT Press.
Best, C. C. & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech, 46,* 183-216.
Blevins, J. (2004). *Evolutionary Phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.
Bonardi, C., Graham, S., Hall, G. & Mitchell, C. (2005). Acquired distinctiveness and equivalence in uman discrimination learning: Evidence for an attentional process. *Psychonomic Bulletin & Review, 12,* 88-92.
Boysson-Bardies, B. de, Schonen, S. de, Jusczyk, P., MacNeilage, P. F. & Morton, J. (eds.). (1993). *Developmental Neurocognition: Speech and face processing in the first year of life.* Dordrecht: Kluwer.
Boysson-Bardies, B. de & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, *67*, 297-319.
Brent, M. R. & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* B33-B44.
Brown, H. & Gaskell, M. G. (2014). The time-course of talker-specificity and lexical competition effects during word learning. *Language, Cognition and Neuroscience, 29,* 1163-1179.
Bybee, J. (2001). *Phonology and Language Use.* Cambridge: Cambridge University Press.
Bybee, J. (2010). *Language, Usage and Cognition.* Cambridge: Cambridge University Press.
Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J. & Witherington, D. (2000). Travel broadens the mind. *Infancy, 1,* 149-219.
Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.
Cristià, A., McGuire, G. L., Seidl, A. & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics, 39,* 388-402.
Dar, M. (2016) Effects of development on cross-language speech perception. Unpublished PhD thesis. University of York.
Davis, B. L. & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research, 38*, 1199-1211.
Davis, B. L. & MacNeilage, P. F. (2000). An embodiment perspective on the acquisition of speech perception. *Phonetica, 57,* 229-241.
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 (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.
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.
DePaolis, R. A., Vihman, M. M. & Keren-Portnoy, T. (2014). When do infants begin recognizing familiar words in sentences? *Journal of Child Language, 41*, 226-239.
Drosopoulos, S., Schulze, C., Fischer, S., & Born, J. (2007). Sleep's function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychology: General, 136,* 169-183.
Dumay, N. & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18,* 35-39.
Dumay, N. & Gaskell, M. G. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition, 123,* 119-132.
Edwards, J., Munson, B. & Beckman, M. E. (2011). Lexicon-phonology relationships and dynamics of early language development – a commentary on Stoel-Gammon's 'Relationships between lexical and phonological development in young children'. *Journal of Child Language, 38*, 35-40.
Eimas, P. D., Siqueland, E. R., Jusczyk, P. W. & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303-306.
Elbers, L. & Wijnen, F. (1992). Effort, production skill, and language learning. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications.* Timonium, MD: York Press.
Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27,* 305-352.
Elsen, H. (1996). Two routes to language. *First Language, 16,* 141-158.
Fagan, M. K. (2009). Mean length of utterance before words and grammar: Longitudinal trends and developmental implications of infant vocalizations. *Journal of Child Language, 36,* 495-527.
Fay, D. & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry, 8,* 505-520.
Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51*, 419-439. Reprinted in Vihman & Keren-Portnoy (2013).
Fernald, A. & Marchman, V. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83,* 203-222.
Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72*, 1003-1015.
Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology, 1*, 5-39.
Foulkes, P. & Docherty, G. J. (2006). The social life of phonetics and phonology, *Journal of Phonetics, 34*,409-438.
Friedrich, M., Wilhelm, I., Born, J. & Friederici, A. (2015). Generalisation of word meanings during infant sleep. *Nature Communications, 6,* 6004 [DOI: 10.1038/ncomms7004].
Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84,* 474-496.

Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R. & Stackhouse, A. (2014). Sleep underpins the plasticity of language production. *Psychological Science, 25*, 1457-1465.
Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1166-1183.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251-279.
Goldstein, M. H., King, A. P. & West, M. J. (2003). Social feedback to infants' babbling facilitates rapid phonological learning. *Proceedings of the National Academy of Sciences, 100,* 8030-8035.
Goldstein, M.H. & Schwade, J. A. (2008). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Psychological Science, 19,* 515-523.
Goudbeek, M., Swingley, D. & Smits, R. (2009). Supervised and unsupervised learning of multidimensional acoustic categories. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1913-1933.
Halle, M. (1971). *The Sound Pattern of Russian*: *A linguistic and acoustical investigation.* Berlin: Walter de Gruyter.
Hallé, P., & Boysson-Bardies, B. de. (1994). Emergence of an early lexicon: Infants' recognition of words. *Infant Behavior and Development, 17,* 119-129.
Hallé, P. & Boysson-Bardies, B. de. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development, 19,* 435-451.
Heitner,R. M. (2004). The cyclical ontogeny of ontology: An integrated developmental account of object and speech categorization. *Philosophical Psychology, 17,* 45-57.
Henderson, L. M., Weighall, A., Brown, H., & Gaskell, M. G. (2013). On-line lexical competition during spoken word recognition and word learning in children and adults. *Child Development, 84*, 1668–1685.
Henderson, L. M., Devine, K., Weighall, A. & Gaskell, G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology, 51,* 406-417.
Herbert, J., Gross, J. & Hayne, H. (2007). Crawling is associated with more flexible memory retrieval by 9-month-old infants. *Developmental Science, 10,* 183-189.
Horváth, K., Myers, K., Foster, R. & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of Sleep Research*. DOI: 10.1111/jsr.12306
Houston, D. M. & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1570-1582.
Houston, D. M. & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 1143-1154.
Huber, R. & Born, J. (2014). Sleep, synaptic connectivity, and hippocampal memory during early development. Trends in Cognitive Sciences, 18, 141- 152. doi.org/10.1016/j.tics.2013.12.005
Huttenlocher, J. (1974). The origins of language comprehension. In R. L. Solso (ed.), *Theories in Cognitive Psychology*. Hillsdale, NJ: Lawrence Erlbaum.

Icht, M. & Mama, Y. (2015). The production effect in memory: A prominent mnemonic in children. *Journal of Child Language, 42,* 1102-1124.

Jabès, A. & Nelson, C. A. (2015). Twenty years after "The ontogeny of human memory: A cognitive neuroscience perspective," where are we? *International Journal of Behavioral Development, 39,* 293-303.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.

Johnson, E. & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44,* 548-567.

Johnson, E. K. & Tyler, M. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*, 339-345.

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

Kamhi, A.G., Catts, H.W. & Davis, M.K. (1984). Management of sentence production demands. *Journal of Speech and Hearing Research, 27*, 329-338.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18,* 111-116.

Keren-Portnoy, T. & Segal, O. (in press). Phonological development in Hebrew-learning infants and toddlers: perception and production. To appear in R. Berman (ed.), *Acquisition of Hebrew from Infancy to Adolescence.* TiLAR. Amsterdam: John Benjamins.

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.

Keren-Portnoy, T., Vihman, M. M. & Lindop Fisher, R. (2015). Isolated words in input to infants: A critical wedge? Presented at Workshop on Infant Language Development, Stockholm.

Kuhl, P. K. (1986). Reflections on infants' perception and representation of speech. In J.S. Perkell & D. H. Klatt, *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831-843.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9,* F13-F21.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). *Philosophical Transactions of the Royal Society, B: Biological Sciences, 363,* 979-1000.

Kumaran, D. & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review, 119,* 573-616.

Lewkowicz, D. & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences, 13,* 470-478.

Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning, 60,* S2, 45-63.

Macken, M. A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua, 49*, 11-49. Reprinted in Vihman & Keren-Portnoy (2013).

MacLeod, C., Gopie, N., Hourihan, K. L., Neary, K. R. & Ozubko, K. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 671-685.

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.

Maurer, D. & Werker, J. F. (2013). Perceptual narrowing during infancy: A comparison of language and faces. *Development Psychobiology, 56,* 154-178.

Maye, J., Werker, J. F. & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–B111.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General, 142*, 1190-1210. doi: 10.1037/a0033812.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C. Rogers, T. T., Seidenberg, M. S. & Smith, L. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences, 14,* 348-356.

McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419-517.

McCune, L. (1992). First words: A dynamic systems view. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press.

McCune, L. (2008). *How Children Learn to Learn Language: Productive skills in a dynamic system.* Oxford: Oxford University Press.

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

McGillion, M. M., Matthews, D., Herbert, J., Pine, J., Vihman, M. M., Keren-Portnoy, T. & DePaolis, R. A. (in press). What paves the way to conventional language? The predictive value of babble, pointing and SES. *Child Development.*

McGregor, K. & Johnson, A. (1997). Trochaic template use in early words and phrases. *Journal of Speech, Hearing and Language Research, 40,* 1220-1231.

Menn, L., Schmidt, E. & Nicholas, B. (2013). Challenges to theories, charges to a model: The Linked-Attractor model of phonological development. In M. M. Vihman & T. Keren-Portnoy (eds.), *The Emergence of Phonology: Whole word approaches, cross-linguistic evidence.* Cambridge: Cambridge University Press.

Menn, L. & Vihman, M. M. (2011). Features in child phonology: inherent, emergent, or artefacts of analysis? In N. Clements & R. Ridouane (eds.), *Where Do Phonological Features Come From? The nature and sources of phonological primitives* (pp. 261-301). Amsterdam: John Benjamins.

Messum, P. & Howard, I. S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than my imitation. *Journal of Phonetics, 53,* 125-140.

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.

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

O'Reilly, R. C. & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advancesin the complementary learning systems framework. *Trends in Coognitive Sciences, 6,* 505-510.

Oviatt, S. (1980). The emerging ability to comprehend language: An experimental approach. *Child Development*, *50*, 97-106.

Pascalis, O., De Haan, M. & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science, 296,* 1321-1323.

Payne, E., Post, B., Astruc, L., Prieto, P. & Vanrell, M. M. (2012). Measuring child rhythm. *Language and Speech, 55*, 203-229.

Pegg, J. E. & Werker, J. F. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America, 102,* 3742-3753.

Piaget, J. (1951). *Play, Dreams and Imitation in Childhood* (translated by M. Cook). New York: Norton.

Piaget, J. (1952). *The Origins of Intelligence* (translated by C. Gattegno & F. M. Hodgson). London: Heinemann.

Piaget, J. (1954). *The Construction of Reality*. New York: Basic Books.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (eds.), *Frequency Effects and the Emergence of Linguistic Structure.* Amsterdam: John Benjamins.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46,* 115-154.

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology, 25*, 143–170.

Priestly, T. M. S. (1977). One idiosyncratic strategy in the acquisition of phonology. *Journal of Child Language, 4*, 45-66. Reprinted in Vihman & Keren-Portnoy (2013).

Redford, M. A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics, 53,* 141-152.

Richmond, J. & Nelson, C. (2007). Accounting for change in declarative memory: A cognitive neuroscience perspective. *Developmental Review, 27,* 349-373.

Richmond, J. & Nelson, C. (2009). Relaitonal memory during infancy: Evidence from eye-tracking. *Developmental Science, 12,* 549-556.

Robinson, A. J. & Pascalis, O. (2004). Development of flexible visual recognition memory in human infants. *Developmental Science, 7,* 527-533.

Ruff, H. A. & Rothbart, M. K. (1996). *Attention in Early Development.* Oxford: Oxford University Press.

Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Schacter, D. & Moskovitch, M. (1984). Infants, amnesics, and dissociable memory systems. In M. Moscovitch (ed.), *Advances in the Study of Communication and Affect* (v. 9, 173-216). New York: Plenum Press.

Scott, L. S., Pascalis, O. & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science, 16,* 197-201.
Segal, O., Hejli-Assi, S. & Kishon-Rabin, L. (2016). The effect of listening experience on the discrimination of /ba/ and /pa/ in Hebrew-learning and Arabic-learning infants. *Infant Behavior and Development, 42,* 86-99.
Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E.E. David, Jr. & P.B. Denes (eds.), *Human Communication: A Unified View*. New York: McGraw-Hill.
Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3-45.
Swingley, D. (2005). Eleven-month-olds' knowledge of how familiar words sound. *Developmental Science, 8,* 432-443.
Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society*, *B, 364*, 3617–3632.
Takashima, A., Bakker, I., van Hell, J. G., Janzen, G. & McQueen, J. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *Neuroimage, 84,* 265-278.
Thelen, E. & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.
Thiessen, E. D. & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- and 9-month-old infants. *Developmental Psychology, 39,* 706-716.
Thiessen E.D., & Saffran, J.R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning & Development, 3,* 73-100.
Tincoff, R. & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10,* 172–175.
Tipper, S. (1992). Selection for action: The role of inhibitory mechanisms. *Current Directions in Psychological Science, 1,* 105-109.
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.
Torkildsen, J. V. K., Hansen, H. F., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I. & Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language, 108,* 73-88.
van de Weijer, J. (1998). *Language input for word discovery*. (v. 9). Nijmegen: Max Planck Institute for Psycholinguistics.
Vihman, M. M. (1981). Phonology and the development of the lexicon: evidence from children's errors. *Journal of Child Language*, *8*, 239-264.
Vihman, M. M. (1992). Early syllables and the construction of phonology. In C.A. Ferguson, L. Menn & C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*. Parkton, MD: York Press.
Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, *21*, 61-82.

Vihman, M. M. (1996). *Phonological Development: The origins of language in the child*. Oxford: Basil Blackwell.
Vihman, M. M. (2014). *Phonological Development: The first two years*. (2nd ed.) Malden, MA: Wiley-Blackwell.
Vihman, M. M. (2016). Prosodic structures and templates in bilingual phonological development. *Bilingualism: Language and Cognition, 19,* 69-88.
Vihman, M. M. (in press). The development of prosodic structure: A usage-based approach. In P. Prieto & N. Esteve-Gilbert, *Prosodic Development in First Language Acquisition.* Trends in language acquisition research (TiLAR). Amsterdan: John Benjamins.
Vihman, M. M. & Croft, W. (2007). Phonological development: Toward a 'radical' templatic phonology. *Linguistics, 45,* 683-725. Reprinted in. Vihman & Keren-Portnoy (2013).
Vihman, M. M., DePaolis, R. A., & Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning, 64: Suppl. 2, 121-140.*
Vihman, M. M. & Keren-Portnoy, T. (eds.) (2013). *The Emergence of Phonology: Whole word approaches, cross-linguistic evidence.* Cambridge: Cambridge University Press.
Vihman, M. M. & Kunnari, S. (2006). The sources of phonological knowledge: A cross-linguistic perspective. *Recherches Linguistiques de Vincennes, 35,* 133-164.
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.
Vihman, M. M. & Majorano, M. (in press). The role of geminates in infants' early words and word-form recognition. *Journal of Child Language.*
Vihman, M. M. & McCune L. (1994). When is a word a word*?* *Journal of Child Language, 21,* 517-542.
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.
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.
Vihman, M. M. & Velleman, S. L. (2000). Phonetics and the origins of phonology. In N. Burton-Roberts, P. Carr & G. Docherty (eds.), *Phonological knowledge: Its nature and status* (pp. 305-339). Oxford: Oxford University Press.
Vihman, M. M., Velleman, S. L. & McCune, L. (1994). How abstract is child phonology? Towards an integration of linguistic and psychological approaches. In M. Yavas (ed.), *First and Second Language Phonology*. San Diego: Singular Publishing.
Vihman, M. M. & Vihman, V-A. (2011). From first words to segments: A case study in phonological development. In I. Arnon & E. V. Clark (eds.), *Experience, Variation, and Generalization: Learning a first language.* (Trends in Language Acquisition Research 7.*)* (pp. 109-133). Amsterdam: John Benjamins.
Vihman, M.M. & Wauquier, S. (in press). Templates in child language. In M. Hickmann, E. Veneziano & H. Jisa (eds.), *Sources of variation in first language acquisition: Languages, contexts, and learners.* Trends in language acquisition research (TiLAR). Amsterdam: John Benjamins.

Walker, M. P. & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron, 44,* 121-133.

Wedel, A. B. (2007). Feedback and regularity in the lexicon. *Phonology, 24*, 147-185.

Werker, J. F. & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language, Learning and Development, 1,* 197-234.

Werker, J. F. & Fennell, C. T. (2004). Listening to sounds versus listening to words: Early steps in word learning. In D. G. Hall & S. Waxman (eds.), *Weaving a Lexicon* (pp. 79-109). Cambridge, MA: MIT Press.

Werker, J. F. & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology, 37,* 278-296.

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49-63.

Werker, J. F., Yeung, H. H. & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science, 21,* 221-226. DOI: 10.1177/0963721412449459

Wilhelm, I., Prehn-Kristensen, A. & Born, J. (2012a). Sleep-dependent memory consolidation - What can be learnt from children? *Neuroscience and Behavioral Reviews, 36,* 1718-1728.

Yeung, H. H. & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition, 113,* 234-243.

Zamuner, T. S., Morin-Lessard, E., Strahm, S. & Page, M. P. A. (in press). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language.*

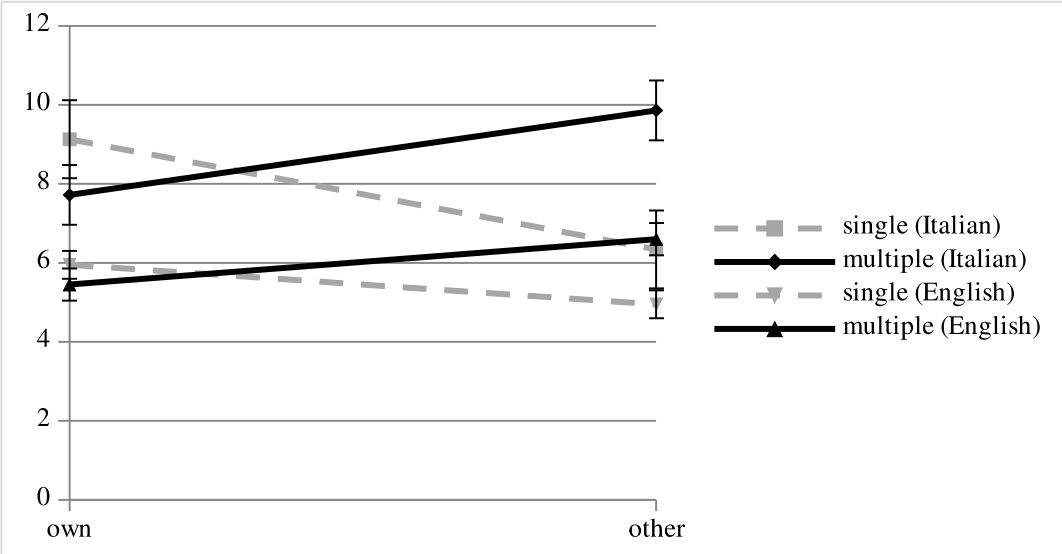
Table 1. First words in four languages (based on observational research studies [English, French] or diary studies [Estonian, German])

GERMAN: Annalena, 8-10 mos. (Elsen, 1996)	
das da 'that one there' /das da/	[data]
Mama 'mama' /mama/	[mama]
Papa 'papa' /papa/	[baba]
pieppiep 'peeppeep' /pipip/	[pɪpɪ]
Teddy /tɛdi/	[dɛdɛ]
ENGLISH: Alice, 9-10 mos. (Vihman, Velleman & McCune, 1994)	
baby	[pɛpɛ:], [tɛɪti:]
daddy	[dæ]
hi	[hɑ:i:], [ʔɑ:jɛ], [hɑɪje] [haɪjʌ]...
mommy	[m̩:ɑn:ə]
no	[njæ̃]
FRENCH: Laurent, 10 mos. (Vihman, 1993)	
allo 'hello' [alo]	[hailo], [ailo], [haljo], [aljo], [alo]
donne (le) 'give (it)' [dʌnlø]	[dlə], [də], [ldɛ], [heldɔ]
l'eau-l'eau 'bottle (nursery word)' [lolo]	[ljoljo]
non [nɔ̃] 'no'	[ne]
tiens [tjɛ̃] 'here, take it'	[ta]
ESTONIAN-ENGLISH: Kaia, 11-15 mos. (Vihman, 2016)	
anna 'give' /an:a/	[an:an:a]
headˈaega 'byebye' /heatˈaeka/	[dada]
kiisu 'kitty' /ki:su/	[ki̥:tɔ̥]
mõmmi 'teddybear' /mɤm:i/	[mʌm]
naba 'belly button' /napa/	[baba]
nämma 'yum' /næm:a/	[mæm:]

Table 2. Phonological templates in later words in four languages. ‘Selected’ words are close to the target, suggesting possible sources for the template; ‘adapted’ words show changes to the target that assimilate it to the template. < > = schematic template form; C = consonant, V = vowel, C_o = optional consonant slot

select		*adapt*	
target word	*child form*	*target word*	*child form*
GERMAN*:* Annalena, 10-12 mos. < σ_1 σ_1 >, i.e., reduplicated syllables (Elsen, 1996)			
Pipi /ˈpipi/ 'peepee'	[pipi:]	*Annalena* /ˈanalena/	[nana]
wauwau /ˈvauvau/ 'bowwow'	[vava]	*kikeriki* /kikeriˈki::/ 'cock-a-doodle-do'	[ki:ki:]
		Bauch /baux/ 'belly'	[baba]
		Tag /tak/ '(good)day'	[dada]
		Zahn(bürste) /ˈtsa:nbürʃtə/ 'tooth(brush)'	[nana]
ENGLISH*:* Alice, 14 mos. <CVCi> (Vihman et al., 1994)			
baby	[bebi]	*bottle*	[baɖi, ba:tʃi, batji]
daddy	[tæɟi]	*hiya*	[ha:ji]
lady	[jɛiji]		
mommy	[maɲi]		
FRENCH*:* Laurent, 15 mos. < C_oVlV > (Vihman & Kunnari, 2006)			
allo ‘hello’ /alo/	[alo]	*canard* ‘duck’ /kanaʁ/	[kɔla]
dans l'eau, de l’eau 'in/some water’ /dɑ̃lo/, /dəlo/	[dəlo]	*chapeau* ‘hat’/ʃapo/	[bɔlo]
ballon ‘big ball’ /balɔ̃/	[palɔ]	*la brosse* ‘the brush’ /labrʌs/	[bəla]
pas là 'not there' /pala/	[pala]	*la cuillère* ‘the spoon’ /lakɥijɛʁ/	[kola]
		voilà 'there you are' /vwala/	[lala]
ESTONIAN: Kaia, 16 mos. < aC:V> (Vihman, 2016)			
anna /ˈan:na/ ‘give'	[an:a]	*õue* /ˈɤʊ:we/ ‘to outside’	[au:a]
		juua /ˈju:wa/ ‘to drink’	[au:a]
		auto /ˈau:to/ ‘car’	[at:o]
		lutti /ˈlut:ti/ ‘pacifier'	[at:i]
		lahti /ˈlah:ti/ ’open, unstuck'	[at:i]

Figure 1. Infants with single VMS (dashed lines) compared with infants with multiple VMS (solid lines) in their response to a passage (British study, DePaolis et al., 2011) or a list of isolated words (Italian study, Majorano et al., 2014), each featuring a particular VMS.


12
10
8
6
4
2
0
own
other
single (Italian)
multiple (Italian)
single (English)
multiple (English)