

The role of mirror neurons in the ontogeny of speech

Marilyn May Vihman
University of Wales Bangor

School of Psychology UWB
Bangor, Gwynedd LL57 2DG
United Kingdom
tel. 44 1248 383 775
fax 44 1248 382 599
email m.vihman@bangor.ac.uk

In Stamenov, Maxim & Victor Gallese. *Mirror Neurons and the evolution of brain and language*. Amsterdam: John Benjamins.

It has been suggested that children's own vocal patterns play a key role in the development of segmental representations of adult words (Vihman, 1991, 1993; Vihman & DePaolis, 2000). The discovery of mirror neurons provides a neurophysiological mechanism for such an 'articulatory filter'. Assuming that only *within repertoire* behaviors can elicit mirror responses, child production of adult-like syllables would be a prerequisite for this kind of matching or filtering. This paper will outline the developmental shift in perception from prosodic to segmental processing over the first year of life and relate that shift to the first major maturational landmark in vocal production, the emergence of canonical syllables. We speculate that it is the activation of the relevant mirror neurons, consequent upon that maturational change, that makes possible the uniquely human shift to segmentally based responses to speech and then to first word production.

Advances in speech perception: from prosodic to segmental patterns

Over the past decade or so experimental work in infant speech perception has increasingly turned from the early focus on infant capacity for discrimination between speech sounds, whether native or non-native, to attempts to probe advances in familiarity with the ambient language which would imply some kind of longer-term representation for speech. Table I provides a summary of those studies. From the division of the table into studies providing evidence of infant knowledge of the *prosody* of speech as opposed to *segmental patterning* it is possible to see a clear developmental trend: For the first six months it is primarily prosodic patterns that underlie a familiarity response to speech. Exceptional evidence of very early (holistic) response to segmental patterns involves stimuli that can be assumed to be imbued with strong affect or 'personal relevance' for infants (e.g., the infant's own name: Mandel, Jusczyk & Pisoni, 1995, and family terms for 'mama' and 'papa': Tincoff & Jusczyk, 1999; for elaboration of the notion of 'personal relevance', see Van Lancker, 1991).

[Insert Table I about here.]

Infants' apparently greater early memory for prosodic patterns follows naturally from the fact that the foetus gains linguistic experience already in the womb, through hearing the sound of the mother's voice both 'internally' and from the outside, as filtered through the amniotic fluid (DeCasper & Fifer, 1980; Querleu, Renard & Versyp, 1981; Hepper, Scott & Shahidullah, 1993). The prosodic information present in the lower frequency bands of the signal can reach foetal ears, once the auditory system is completely formed (by the final trimester of pregnancy), while segmental information, much of which is carried by higher frequencies, cannot. In a series of studies Fernald has shown that prosody must indeed provide the initial entry into language, due not only to its preestablishment as an acoustic signal before birth but also to the intrinsic affective links between particular prosodic patterns and communicative meanings (Fernald, 1989, 1992).

With regard to segmental patterning we first see attention to the child's own name (as early as 4 months) and association of parent terms with the appropriate parent (by 6 months), as noted above. Aside from these exceptional word forms, preferential attention to a (trained) segmental pattern is not reported until 7.5 months of age (Jusczyk & Aslin, 1995). Thereafter, steady gains in attention to segmental patterning can be seen (e.g., preference for native language phonotactics: Friederici & Wessels, 1993, and a narrowing of attention to consonantal contrasts from broadly 'universal' to native-language only: Werker & Tees, 1984; Best, 1994). To complete this picture, Myers, Jusczyk, Kemler Nelson, Charles-Luce, Woodward & Hirsh-Pacek (1996) showed that at 11 months, but not at 4.5 or even 9 months, infants

looked longer to ‘coincident’ passages, which had brief pauses inserted only *between* words, than to ‘non-coincident’ passages, which included two- or three-syllable words *interrupted* by such pauses. Unlike earlier studies which demonstrated greater attention to coherent vs. interrupted units (whether clauses or phrases) even when the stimuli were low-pass filtered to remove segmental information (Kemler Nelson, Hirsh-Pasek, Jusczyk & Wright Cassidy, 1989; Jusczyk & Kemler Nelson, 1996; Jusczyk et al., 1992), the coherent-word effect was not obtained when only prosodic information was available.

First perceptual representations of speech forms

Despite intensive experimental work on infant responses to speech for over twenty years, Hallé and Boysson-Bardies (1994) was the first study to examine infant responses to *untrained* speech forms. These authors tested 11-month-old French infants on word patterns expected to be familiar from everyday exposure. The words were chosen from those produced early in the second year by French infants in an earlier study and were matched with phonetically similar words of low frequency in the adult lexicon. Exposed to test lists of 12 words of each kind in the head-turn preference technique, infants were found to attend longer to the ‘familiar’ than to the ‘rare’ words. In a follow up study, Hallé and Boysson-Bardies (1996) explored the phonetic basis for the familiarity effect by removing the initial consonant (the familiarity effect was eliminated), changing voicing or manner of the initial consonant (the familiarity effect was observed), or changing manner of the second consonant (the effect just failed to reach significance).

Since the accentual pattern of French (iambic, or weak-strong, based primarily on lengthening of word- or phrase-final syllables) is the opposite of the dominant stress pattern found in English content words (trochaic, or strong-weak), we sought to replicate the French results with infants exposed to British English (Vihman, Nakai & DePaolis, 2000). In a base-line experiment we used lists including seven trochaic words and five iambic words or phrases (e.g., ‘familiar’ *apple, baby, a ball, fall down*, vs. ‘rare’ *bridle, maiden, a bine, taboo*), the familiar words taken from previous studies of early word production in children acquiring English. While 11-month-olds were found to attend longer to the familiar words ($p < .05$), 9-month olds failed to show a significant difference between listening times to the two lists.

In a second experiment we sought to establish the role of prosody in the familiarity effect for English-learning children by contrasting lists with the unaltered familiar words versus the same familiar words under altered stress (e.g., *baby > baby*). No significant difference was found, suggesting that the stress pattern did not constitute an essential part of the infants’ lexical representations. We validated this finding in an additional experiment in which we contrasted the list of altered familiar words used in the previous experiment with a list of altered *unfamiliar* words; under these conditions the former received significantly longer looks ($p < .001$), again demonstrating that the infants could ‘listen through’ the stress pattern to recognize familiar words.

A final pair of experiments was designed to test whether accented syllables are more fully specified in infant word representations than unaccented syllables. We predicted that there would be a difference between infants’ representation of early French words, which are iambic, and early English words, which are mainly trochaic. We therefore presented English-learning infants with (trochaic) lists of rare words in

contrast with familiar words with, first, a change in manner of articulation affecting all medial consonants (e.g., *bubbles* > *bumbles*, *piggy* > *pingy*) or, in a second experiment, all initial consonants (*mubbles*, *figgy*). As anticipated, the infants attended longer to the familiar words despite the change in the second consonant ($p < .01$) but failed to show significantly longer looking times to the familiar words with changed initial consonants, suggesting that the familiar words were no longer recognizable in that condition. We concluded that the accentual pattern of the adult language may influence which details are noted in early word representations, changes to the initial syllable blocking word recognition in English but not in French. In summary, these experiments suggest that the first representations for words and phrases are influenced by prosodic patterning (hence the ambient-language differences in infant responses to changes in initial vs. second syllable onset-consonant) yet by 11 months prosody itself constitutes a less essential property of these representations than segmental patterning (as found in our stress-change experiment).

The articulatory filter and mirror neurons

A plausible source for the shift from a largely prosodic to a primarily segmental basis for attention to speech patterns can be found in the developmental milestones for production in the first year. In hearing infants canonical babbling, or the rhythmic production of consonant-vowel (CV) sequences with adult-like timing, is reliably reported by parents and confirmed by laboratory recordings to occur at about 6-8 months of age (Oller, 1980; Stark, 1980; Lindblom & Zetterstrom, 1986). The coincidence in timing of the production milestone with the shift to first perceptual responses suggestive of segmental representation is striking. Unfortunately, however, infant speech perception studies to date have included neither canonical babbling status for infant participants nor individual perceptual responses, so we lack so much as a correlational study showing the onset of CV production in relation to changes in attention to segmental speech patterns in perception.

The emergence of easily recognized babbled syllables with adult-like timing in the middle of the first year appears to be maturationally based and fits into a broader framework of rhythmic motoric advances that occur around that age (Thelen, 1981). One interpretation of the developmental match between the shift to attention to segments and the onset of CV production is the articulatory filter hypothesis (Vihman, 1993). On this account, the experience of frequently producing CV syllables sensitizes infants to similar patterns in the input speech stream (note that deaf infants fail to persevere in CV syllable production at the typical age: Oller & Eilers, 1988). As in the ‘cocktail party’ effect produced in adults when their own name occurs in an unattended conversation (Wood & Cowan, 1995), particular segmental patterns would now begin to ‘pop out’ of input speech which previously might have constituted only ‘background music’ for the infant listener. Infants could be expected to differ in their sensitivity to the putative ‘match’ of own production patterns to adult input; presumably, the process would not be instantaneous but cumulative, leading eventually to the best-represented adult patterns – those closely resembling the child’s own most typical production patterns – forming the basis for first words in expressive infants. Most children could be expected to show some influence of their own incipient adultlike syllable production on their attention to speech, a proposition currently being tested in our lab.

The mirror neuron findings provide unanticipated neurophysiological support for this speculative idea. In the course of making single-cell recordings of the premotor cortex in monkeys di Pellegrino, Fadiga, Fogassi, Gallese & Rizzolatti (1992) discovered that ‘when the monkey observes a motor action that is *present in its natural movement repertoire*, this action is automatically covertly retrieved’ (Fadiga, Fogassi, Pavesi &

Rizzolatti, 1995, p. 2608; my emphasis). Fadiga et al. (1995) provide indirect neurological evidence that ‘in humans [too] there is a neural system matching action observation and execution...The observation of an action automatically recruits neurons that would normally be active when the subject executes that action’ (p. 2609). Practice in performing a particular motor routine (e.g., producing CV syllables) lays the groundwork for the activation of the same motor neurons when similar routines (e.g., adult word forms similar to the infant’s babbling patterns) are *produced by others*.

This account provides a natural mechanism for imitation of within-repertoire motor behaviors. Thus, some (but not all) 4-5 month-old infants in an auditory/visual matching experiment involving isolated vowels spontaneously imitated the vowels (Kuhl & Meltzoff, 1988). Kuhl and Meltzoff proposed that “infants make an intramodal auditory-auditory match; and second, they develop a set of auditory-articulatory mapping rules” (p. 254). The mirror neurons provide a more direct neuromotor mechanism for effecting imitation as a by-product of attending to vowels at 4-5 months. The later emergence of speech-like CV patterns in an infant’s vocal repertoire would then provide the basis for the more sophisticated capacity to pay privileged attention to and represent or remember the particular speech forms of the ambient language. The critical point is the requirement that a movement be present in the individual’s natural repertoire before the mirror system can effect a match of observed to potential action patterns. By this account, it is only after the individual infant’s neurological system has been prepared by the onset of rhythmic babbling that the mirror system relating heard patterns to potential production patterns can begin to function to highlight a subset of the patterns embedded in the fast-changing input speech signal.

First word production as the product of an articulatory filter

It has long been recognized that the first words tend to be relatively accurate, arguably due to selection on phonological grounds (Ferguson & Farwell, 1975; Schwartz, 1988), and that they resemble babbling patterns, both generally and for individual children (Vihman, Macken, Miller, Simmons & Miller, 1985). Analyses of later word forms indicate that a given child’s first well-practiced, consistent supraglottal production patterns (‘vocal motor schemes’: McCune & Vihman, 1987) provide the basis for the later development of ‘word templates’ (Vihman & Velleman, 2000). These word templates abstract from and extend the piecemeal learning evidenced by the selection patterns of first words (Table II).

[Insert Table II about here.]

The apparent paradox of such early word selection – how do children know which words not to attempt, or which sounds they cannot yet produce? (Stemberger & Bernhardt, 1999) – does not arise if we assume that the first words result from infant matching of own vocal patterns to the input speech signal. The evidence from neurophysiology that a mirror system may mediate perception-action links, imitation, and learning, although still speculative, places the notion of an articulatory filter in the first year of life on firmer ground. More conclusive evidence will have to come from ongoing direct empirical research with infants.

References

- Best, Catherine T. (1994). The emergence of language-specific phonemic influences in infant speech perception. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception* (167-234). Cambridge, MA: MIT Press.
- Cooper, Robin P. & Richard N. Aslin. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584-1595.
- DeCasper, Anthony J. & William P. Fifer. (1980). Of human bonding. *Science*, *208*, 1174-1176.
- di Pellegrino, Giuseppe, Luciano Fadiga, Luciano Fogassi, Vittorio Gallese & Giacomo Rizzolatti (1992). Understanding motor events. *Experimental Brain Research*, *91*, 176-180.
- Fadiga, Luciano, Luciano Fogassi, G. Pavesi & Giacomo Rizzolatti. (1995). Motor facilitation during action observation. *Journal of Neurophysiology*, *73*, 2608-2609.
- Ferguson, Charles A. & Carol B. Farwell (1975). Words and sounds in early language acquisition. *Language*, *51*, 419-439.
- Fernald, Anne. (1989). Intonation and communicative intent in mothers' speech to infants. *Child Development*, *60*, 1497-1510.
- Fernald, Anne. (1992). Human maternal vocalizations to infants as biologically relevant signals. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind* (391-428). Oxford: Oxford University Press.
- Friederici, Angela D. & Jeanine M. I Wessels. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*, 287-295.
- Hallé, Pierre & Bénédicte de Boysson-Bardies. (1994). Emergence of an early lexicon. *Infant Behavior and Development*, *17*, 119-129.
- Hallé, Pierre & Bénédicte de Boysson-Bardies. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, *19*, 435-451.
- Hepper, P. G., D. Scott & S. Shahidullah. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, *11*, 147-153.
- Jusczyk, Peter W. & Richard N. Aslin. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1-23.
- Jusczyk, Peter W., Anne Cutler & Nancy J. Redanz (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*, 675-687.
- Jusczyk, Peter W. & Deborah G. Kemler Nelson. (1996). Syntactic units, prosody, and psychological reality during infancy. In J. L. Morgan & K. D. Demuth (Eds.), *Signal to syntax* (389-408). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Jusczyk, Peter W., Deborah G. Kemler Nelson, Kathryn Hirsh-Pasek, Lori J. Kennedy, Amanda Woodward & Julie Piwoz (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*, 252-293.
- Kemler Nelson, Deborah G., Kathy Hirsh-Pasek, Peter W. Jusczyk, & Kimberly Wright Cassidy (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*, 55-68.
- Kuhl, Patricia K. & Andrew N. Meltzoff. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy. The Minnesota Symposia on Child Psychology*, *20* (235-266). Hillsdale, NJ: Lawrence Erlbaum.
- Lindblom, Björn & Ralph Zetterström (Eds.), *Precursors of early speech..* Basingstoke, Hampshire: Macmillan Press.

- Mandel, Denise R., Peter W. Jusczyk & David B. Pisoni. (1995). Infants' recognition of the sound pattern of their own names. *Psychological Science*, 6, 315-318.
- McCune, Lorraine & Marilyn M. Vihman. (1987). Vocal motor schemes. *Papers and Reports on Child Language Development*, 26, 72-79.
- Mehler, Jacques, Peter W. Jusczyk, Ghislaine Lambertz, Nilofar Halsted, Josiane Bertoni, & Claudine Amiel-Tison. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.
- Moon, Christine, Robin Panneton-Cooper & William P. Fifer. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16, 495-500.
- Myers, James, Peter W. Jusczyk, Deborah G. Kemler Nelson, Jan Charles-Luce, Amanda L. Woodward & Kathryn Hirsh-Pasek. (1996). Infants' sensitivity to word boundaries in fluent speech. *Journal of Child Language*, 23, 1-30.
- Oller, D. Kimbrough. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), *Child Phonology, I: Production* (93-112). New York: Academic Press.
- Oller, D. K Kimbrough & Rebecca E. Eilers. (1988). The role of audition in infant babbling. *Child Development*, 59, 441-449.
- Querleu, D., X. Renard, F. Versyp, L. Paris-Delrue & G. Crépin. (1988). Fetal hearing. *European Journal of Obstetrics and Reproductive Biology*, 29, 191-212.
- Schwartz, Richard G. (1988). Phonological factors in early lexical acquisition. In M. D. Smith & J. L. Locke (Eds.), *The emergent lexicon* (185-222). New York: Academic Press.
- Stark, Rachel E. (1980). Stages of speech development in the first year of life. In G. Yeni-komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), *Child Phonology*, vol. 1: *Production* (73-92). New York: Academic Press.
- Stemberger, Joseph P. & Barbara H. Bernhardt. (1999). The emergence of faithfulness. In B. MacWhinney (Ed.) *The emergence of language* (417-446). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Thelen, Esther. (1981). Rhythmical behavior in infancy. *Developmental Psychology*, 17, 237-257.
- Tincoff, Ruth & Peter W. Jusczyk. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, 10, 172-175.
- Van Lancker, Diane. (1991). Personal relevance and the human right hemisphere. *Brain and Cognition*, 17, 64-92.
- Vihman, Marilyn M. (1991). Ontogeny of phonetic gestures. In I.G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (69-84). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Vihman, Marilyn M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21, 61-82.
- Vihman, Marilyn M. & Rory A. DePaolis. (2000). The role of mimesis in infant language development. In C. Knight, J. Hurford, & M. Studdert Kennedy (Eds.), *The evolutionary emergence of language*. (000-000). Cambridge: Cambridge University Press.
- Vihman, Marilyn M., Marlys A. Macken, Ruth Miller, Hazel Simmons & James Miller. (1985). From babbling to speech. *Language*, 61, 395-443.
- Vihman, Marilyn M., Satsuki Nakai & Rory A. DePaolis. (2000). The role of accentual pattern in early lexical representation. Poster presented at 12th International Conference on Infant Studies, Brighton, UK. July.
- Vihman, Marilyn M. & Shelley L. Velleman. (2000). The construction of a first phonology. *Phonetica*, 57, 255-266.

- Werker, Janet F. & Robert C. Tees. (1984). Cross-language speech perception. *Infant Behavior and Development*, 7, 49-63.
- Wood, Noelle & Nelson Cowan. (1995). The cocktail party phenomenon revisited. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 255-260.

Table I. Advances in perception and representation of the native language in the first year

Child attends more to...		
	Prosodic patterns	Segmental patterns
At birth	... native language (vs. prosodically dissimilar other language) (Moon et al., 1993)	
By 1 mo	... infant-directed (ID) prosody (vs. adult-directed prosody) (Cooper & Aslin, 1990)	
By 2 mos	... native-language narrative passages and short utterances (vs. prosodically dissimilar other language) (Mehler et al., 1988)	
By 4 mos	... 'coincident' clauses (vs. non-coincident clauses (Jusczyk & Kemler Nelson, 1996) – but not phrases or words (Jusczyk et al., 1992; Myers et al., 1996)	... own name (vs. other name: Mandel, Jusczyk & Pisoni, 1995)
By 6 mos	... word list in native language (vs. prosodically dissimilar language, even when low-pass filtered) (Jusczyk et al., 1993)	... family words for <i>mama</i> and <i>papa</i> , with matching video (Tincoff & Jusczyk, 1999)
Between 6 and 8 months	Emergence of canonical (CV) syllables in production, with adult-like timing (Oller, 1980; Stark, 1980; Lindblom & Zetterström, 1986)	
By 8 mos		... monosyllabic word forms previously trained through narrative passages or word lists (Jusczyk & Aslin, 1995)
By 9 mos.	... native language stress pattern (even when low-pass filtered) (Jusczyk et al., 1993) ... uninterrupted phrases [even when low-pass filtered] (Jusczyk et al., 1992), but not	... native language phonotactics (Friederici & Wessels, 1993)

	words (Myers et al., 1996)	
By 10 mos		[Fail to discriminate non-native consonant contrasts (Werker & Tees, 1984)]
By 11 mos.		<p>... word forms familiar from everyday experience (Hallé & Boysson-Bardies, 1994)</p> <p>... uninterrupted words [but not when low-pass filtered] (Myers et al., 1996)</p> <p>...familiar words [even with a <i>reversal</i> of the accentual pattern – but <i>not</i> with change to the onset C of the accented syllable] (Vihman et al., 2000)</p>

Table II. Relationships between adult and child word forms (Alice)

9-10 Months		
Adult Target	Selected child forms	
<i>baby</i> /beɪbi/	[pɛpɛ:]	
<i>daddy</i> /dædi/	[dæ]	
<i>hi</i> /haɪ/	[ha:i]	
<i>mommy</i> /mami/	[m:an:ə]	
<i>no</i> /nɒ/	[njæ]	
14 Months Template schema <CVCi>		
Adult Target	Selected child forms	Adapted child forms
<i>baby</i> /beɪbi/	[bebi]	
<i>bottle</i> /bɒtəl/		[bɑ:ji]
<i>daddy</i> /dædi/	[tæ:ji]	
<i>hiya</i> /haɪ/		[ha:ji]
<i>lady</i> /leɪdi/	[je:ji]	
<i>mommy</i> /mami/	[ma:ɹi}	

Alice's first spontaneous words, recorded at 9-10 months, are listed here in full. The adult targets for these first words suggest 'selection on phonological grounds': Note that three of the five are disyllables including a single repeated stop or nasal and ending in the vowel /i/. In contrast, some of the words that Alice produced at 14 months fit the template schema she has now evolved ('selected') while others are adapted to fit the schema.