Infant vocalizations in response to speech: Vocal imitation and developmental change

Patricia K. Kuhl[a)]
Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98195

Andrew N. Meltzoff
Department of Psychology, University of Washington, Seattle, Washington 98195

(Received 30 December 1993; revised 3 May 1996; accepted 8 May 1996)

Infants' development of speech begins with a language-universal pattern of production that eventually becomes language specific. One mechanism contributing to this change is vocal imitation. The present study was undertaken to examine developmental change in infants' vocalizations in response to adults' vowels at 12, 16, and 20 weeks of age and test for vocal imitation. Two methodological aspects of the experiment are noteworthy: (a) three different vowel stimuli (/a/, /i/, and /u/) were videotaped and presented to infants by machine so that the adult model could not artifactually influence infant utterances, and (b) infants' vocalizations were analyzed both physically, using computerized spectrographic techniques, and perceptually by trained phoneticians who transcribed the utterances. The spectrographic analyses revealed a developmental change in the production of vowels. Infants' vowel categories become more separated in vowel space from 12 to 20 weeks of age. Moreover, vocal imitation was documented. Infants listening to a particular vowel produced vocalizations resembling that vowel. A hypothesis is advanced extending Kuhl's native language magnet (NLM) model to encompass infants' speech production. It is hypothesized that infants listening to ambient language store perceptually derived *representations* of the speech sounds they hear which in turn serve as targets for the production of speech utterances. NLM unifies previous findings on the effects of ambient language experience on infants' speech perception and the findings reported here that short-term laboratory experience with speech is sufficient to influence infants' speech production.

PACS numbers: 43.71.An, 43.71.Ft, 43.71.Es, 43.70.Bk [RAF]

INTRODUCTION

Speech-production development during the first 2 years of life has been described as a set of universal stages (Kent, 1992; Oller and Lynch, 1992; Stoel-Gammon, 1992). Wide consensus now exists among investigators on specific, orderly changes that occur in the vocalizations produced by young infants in American English (Oller, 1978; Stark, 1980; Stoel-Gammon and Cooper, 1984; Vihman and Miller, 1988) and other languages (Holmgren *et al.*, 1986; Koopmans-van Beinum and van der Stelt, 1986; Roug *et al.*, 1989). Five stages in vocal development can be identified: *reflexive phonation* (0–2 months), in which vegetative or reflexive sounds such as coughing, sneezing, and crying predominate; *cooing* (1–4 months), in which infants produce quasivocalic sounds that resemble vowels; *expansion* (3–8 months), characterized by the occurrence of clear vowels that are fully resonant and a wide variety of new sounds such as yells, screams, whispers, and raspberries; *canonical babbling* (5–10 months), during which infants produce strings of consonant-vowel syllables, such as ''bababa'' or ''mamama,'' and *meaningful speech* (10–18 months), wherein infants mix both babbling and meaningful speech to produce long intonated utterances.

Although there is consensus on describing speech production stages, little is known about the processes by which change in infants' vocalizations are induced. Two factors are critical in the early phases, anatomical change and vocal learning. The young infant's vocal tract is very different from that of the adult, more closely resembling that of a nonhuman primate than that of an adult human (Bosma, 1975; Kent, 1981; Lieberman *et al.*, 1972). The infant's vocal tract is not only much smaller than that of the adult's, it has a broader oral cavity, a tongue mass that is proportionally larger and more anterior, and a more gradually sloping oropharyngeal tract (see Kent, 1992 for review). During the first half year of life the vocal tract undergoes dramatic change as it develops into one that more closely resembles that of the adult human (Sasaki *et al.*, 1977). This anatomical restructuring contributes to increased motor dexterity of the articulators and an increase in the formant range that can be produced. Anatomical changes contribute, at least in part, to the stage-like changes seen in infants' vocalizations worldwide.

A second factor responsible for change in infants' vocalizations, one we know much less about, is *vocal learning*. Human infants listen to ambient language spontaneously and attempt to produce sound patterns that match what they hear. In other words, infants acquire the specific inventory of phonetic units, words, and prosodic features employed by a *particular* language in part through imitation. At the endpoint of infancy, toddlers ''sound like'' a native speaker of their language. *Homo sapiens* is the only mammal that displays vocal learning, the tendency to acquire the species-typical vocal

[a)]Author to whom correspondence and requests for reprints should be addressed.

repertoire by hearing the vocalizations of adults and mimicking them. Humans share this ability with a few select avian species, the songbirds (Konishi, 1989; Marler, 1974), who learn their species-specific songs only if they are auditorially exposed to them during a sensitive period early in life (Nottebohm, 1975).

There is evidence that the experience gained from hearing oneself and others, not solely anatomical change, contributes to speech-production development. Deaf infants' vocalizations differ from those of normal infants. The onset of canonical babbling is delayed in deaf infants, and when it occurs, the babbled utterances differ in duration and timing (Kent *et al.*, 1987; Oller and Eilers, 1988; Oller *et al.*, 1985). Moreover, the phonetic inventories of deaf infants differ from those of normal infants. Deaf infants rely on sounds that are visually prominent, such as /ba/ and /ma/, to a greater extent than normal infants (Stark, 1983; Stoel-Gammon, 1988; Stoel-Gammon and Otomo, 1986). The role of audition in the learning of vocalization is also suggested by the fact that adult speakers of the language speak the dialect produced locally. Hearing a specific language early in life puts an indelible mark on one's speech. Thus, the acquisition of speech is both anatomically constrained and auditorially guided (for an extended discussion of visual and other multimodal influences, see Kuhl and Meltzoff, 1988; Locke, 1993; Meltzoff and Kuhl, 1994).

Identifying the age at which infants begin mimicking the sound patterns they hear is important for theory construction. Relevant data can be adduced from the age at which infants from different language environments begin to produce sounds that are unique to their own native language. Cross-cultural studies in which the sounds produced by young infants were phonetically transcribed suggest that infants' earliest vocalizations do not show an effect of language environment (Holmgren *et al.*, 1986; Koopmans-van Beinum and van der Stelt, 1986; Oller, 1978; Roug *et al.*, 1989; Stark, 1980; Stoel-Gammon and Cooper, 1984; Vihman and Miller, 1988). However, between 10 and 12 months of age, a few studies suggest that infants from different linguistic environments have begun to exhibit differences in their vocalizations (de Boysson-Bardies *et al.*, 1989; de Boysson-Bardies *et al.*, 1984; de Boysson-Bardies *et al.*, 1992). Between 2 and 3 years of age infants from different cultures show clear differences, even in subtle measures (Stoel-Gammon *et al.*, 1994).

One question is whether vocal learning begins only after the first or second year (when language-specific effects are well documented in spontaneous utterances) or whether there is learning taking place earlier. Vocal imitation provides evidence of the potential for learning, and this ability may be present well before the measurement of a corpus of spontaneously produced utterances shows that learning has taken place.

The importance of documenting the development of vocal imitation for theories of speech and language has been discussed by Kent and Forner (1979), Kuhl and Meltzoff (1982, 1988, in press; Meltzoff and Kuhl, 1994), Locke (1993), and Studdert-Kennedy (1986, 1993). Vocal imitation requires that infants recognize the relationship between articulatory movements and sound. In adults the information specifying auditory–articulatory relations is exquisitely detailed (e.g., Perkell *et al.*, 1993). It is as though adults have an internalized auditory-articulatory "map" that specifies the relations between mouth movements and sound. When do infants acquire the auditory-articulatory map?

Experimental studies of vocal imitation in the first year of life are rare. However, intriguing observations suggesting a capacity for vocal imitation are abundant. From Piaget (1962) on, reports have appeared of the imitation of some aspect of speech, typically the prosodic characteristic of pitch (Kessen *et al.*, 1979; Lieberman, 1984; Papousek and Papousek, 1981). In some cases, both the imitation of the prosodic aspects of pitch and vowel formants is suggested (Lieberman, 1984, 1991). The studies have methodological problems that prevent strong inferences about vocal imitation, however. All but one (Kessen *et al.*, 1979) involved natural interactions between infants and adults, and as such are subject to a variety of problems, the most important of which is the question, "Who is imitating whom?" In the Kessen *et al.* (1979) study, infants were reported to match the absolute pitch produced by a pitch pipe. However, infants were tested in multiple sessions over several months, and the issue of whether infants' responses were due to specific training/shaping by the experimenters was unresolved. In more recent work (Legerstee, 1990), utterances were coded by people who were not phonetically trained and no instrumental analysis of infants' vocalizations was provided.

Suggestive evidence for vocal imitation was also reported by Kuhl and Meltzoff (1982). In that study infants were presented with an auditory-visual cross-modal matching task using the vowels /a/ and /i/. Infants viewed two filmed images of a female talker producing the two vowels side by side and heard either /a/ or /i/. The results showed that infants looked longer at the face matching the sound they heard, demonstrating cross-modal matching (Kuhl and Meltzoff, 1982, 1984). Infants also vocalized in response to the adult female's productions, mimicking the intonational pattern they heard. It was shown that infants responded differentially to speech versus nonspeech stimuli, producing significantly more speechlike vocalizations when listening to speech as opposed to nonspeech (Kuhl and Meltzoff, 1982, 1988).

The present experiment examined infants' vocalizations in response to vowels at 12, 16, and 20 weeks of age. The three specific aims were to: (a) compare the acoustic nature of infants' and adults' vocalizations, (b) examine developmental change in the infants' vocalizations between the ages of 12 and 20 weeks of age, and (c) assess vocal imitation in a laboratory setting using the vowels /a/, /i/, and /u/ (as in the words "hop," "heap," and "hoop"). Both perceptual (phonetic transcription) and instrumental (spectrographic) analysis methods were used. This is the first laboratory study on infant vocal imitation in which the model's utterances were computer controlled and both perceptual and instrumental techniques were used to analyze infants' responses to modeled utterances.