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.

I. METHOD

A. Design

A method was developed for examining vocal imitation in young infants. There were three noteworthy aspects of the method. First, the vocal stimulus was presented by a machine. This allowed the stimulus to be quantified completely, repeated in an identical way for each infant, and avoided interchanges that are open to the question, "Who is imitating whom?" Second, Meltzoff and Moore described the necessary logic and controls for investigating imitation in the gestural domain, and their "cross-target" design was adapted for use in vocal imitation (Meltzoff and Moore, 1983a, b). In this design, different speech sounds (e.g., phoneme No. 1, No. 2, No. 3) are presented to groups of infants and used as controls for one another. Because the vocal stimuli are produced by the same talker, at the same distance, at the same rate, and for the same temporal response period, any differential response by infants is attributable to the variable being manipulated, namely, the content of the model's vocalization. If infants' vocalizations differ as a function of stimulus condition and match the model's vocalizations, this suggests vocal imitation. Third, both perceptual (phonetic transcription) and physical (computerized spectrographic) analyses were used. Phonetic transcription is needed to establish whether an infant's response perceptually matches that of the adult phonetic category. Instrumental analyses establish which acoustic dimensions infants are capable of varying and establish how infants manipulate those dimensions when producing sounds that are perceived as different. Using both approaches is important in gaining an understanding of infant vocal development and in demonstrating vocal imitation.

B. Subjects

The subjects were 72 infants, 24 in each of three age groups, 12-, 16-, and 20-weeks old. Infants visited the laboratory for three test sessions, typically on consecutive days. The mean age of the three groups at the initial test was 12.2 weeks (range: 11.57–12.35), 16.1 weeks (range: 15.55–16.49), and 20.3 weeks (range: 19.80–20.32). Approximately equal numbers of girls and boys were tested at each of the three ages. Participation in the experiment was elicited by a letter that was sent to the parents of newborns in the Seattle area. Interested parents returned a postcard that provided details regarding birth and family medical history. Preestablished criteria for inclusion in the study were that infants had no known visual or auditory deficits, had uncomplicated deliveries and were developing normally, were full term (>37 weeks gestational age), normal birthweight (2.5–4.5 kg), and that members of their immediate families had no history of hearing loss. An additional 45 infants failed to complete testing due to crying (8), equipment failure (2), sessions that were more than one week apart (1), or failure to return for all three sessions (34). Parents were paid $25 for completing the experiment.

C. Stimuli

Three vowels served as stimuli: /a/ as in "hop," /i/ as in "heap," and /u/ as in "hoop." The stimuli were presented to infants as auditory-visual face-voice stimuli. It was felt that both the face and the voice would be necessary to create a sufficiently natural situation to induce the infant to produce speech, given that the stimulus was not presented live in a natural setting but in a controlled laboratory setting by a machine.

All vowels were produced by the same female talker. Vowels were audio and video recorded in a studio using a color camera (Ikegami ITE 730A) and high quality microphone (Sony EMC 50 electret), which fed a 3/4 in. videotape machine (JVC CR-8250U). The video focused on a closeup of the woman's face against a black background. The talker, cued by a timer that delivered a tone to her right ear, produced a vowel once every 5 s, attempting to articulate them slowly and clearly. Rather than use a single production of each of the three vowels, we used eight different productions to represent each of the vowel categories on the final stimulus tape. Different exemplars of each category were used to provide some natural variation in the characteristics that do not affect the identification of the three vowels, such as intensity, duration, and fundamental frequency. This variation gave the sense that the set of stimuli were produced in a natural style, and this made them more interesting to listen to and look at. The video tokens were selected such that the series of eight visual stimuli appeared natural but contained no eye blinks during the production of any of the vowels. These eight visual stimuli were edited with minimum loss of auditory or visual fidelity (Convergence Corporation ECS-90) to produce a 10-min videotape for each of the three vowels which contained 16 blocks of the sequence of eight stimuli.

The auditory signals that were used were not recorded with the original videotaped utterances, but corresponded in duration to the sequence of eight video stimuli (each audio-video pair in the sequence had to be within 30 ms in total duration). The sequence of eight auditory vowel stimuli were judged to be excellent exemplars of each of the three speech categories. The stimuli were low-pass filtered at 10 kHz and digitized at a 20-kHz sampling rate using a PDP 11/73 computer. The audio tokens were then dubbed, in sequences of eight stimuli, onto the video stimuli contained on the 10-min videotapes. When dubbing the audio tokens onto the video tokens, synchrony was determined by using the original audio signals that had been recorded on one channel of the videotape to cue the computer to output the new audio tokens. This resulted in audio-video synchrony for the experimental tokens that matched that of the original auditory-visual signals.

Spectrographic measurements confirmed the fact that the first two formant frequencies of the sets of /a/, /i/, and /u/ vowels varied in the predicted fashion. The formant frequency values for the three vowels were as follows: for /a/, $F1$ (M=764.5, range=715–786 Hz), $F2$ (M=1010.3, range 967–1029 Hz); for /i/, $F1$ (M=296.1, range, 285–310 Hz), $F2$ (M=2726.4, range, 2674–2818 Hz); for /u/, $F1$ (M=294.1, range, 280–309 Hz), $F2$ (M=943.4, range, 873–

1013 Hz). These values fall within the range that has previously been reported for adult females' productions of the vowels /a/, /i/, and /u/ (Peterson and Barney, 1952). The average duration of the vowels was 1.63 s (range=1.40–1.87). The vowels were presented at an average intensity of 68-dB SPL (range=65–70 dB) measured at the location of the infant's head (Bruel & Kjaer, A scale, fast). All of the vowels were produced with a rise–fall intonation contour.

D. Equipment and test apparatus

The test suite consisted of two soundproof rooms separated by clear glass. The control room housed a PDP 11/73 computer, video playback machine (JVC CR-8250U) that reproduced the stimulus for infants, a video recorder with high-definition specifications (Sony SL 5600) for recording infants' visual and auditory vocalizations, and an audio cassette recorder (Tascam 122 MKII) for an additional recording of both the infants' and the models' vocalizations.

In the test room the video image was displayed on a 12-in. television screen (NEC JC1215MA) placed on a stand at eye level for the infant. The television screen was embedded within a 1.5-m-high×1.2-m-wide gray cloth-covered panel; the screen was visible through a window cut in the panel. The face was lifesize (21-cm long, 15-cm wide). Two adjacent panels, 1.5-m high×0.9-m wide, also covered in gray cloth, prevented the infant from being distracted by surroundings in the test room. A camera (Panasonic, model WV-135A) was positioned behind a small hole in the front panel. It recorded a closeup of the infant's face (from about 2.5 cm below the chin to top of the forehead). A high-quality microphone (Audio Technica Shotgun, model AT815A) was used to record infants' vocalizations; its cylindrical shape allowed it to protrude through the panel to a position several inches from the infant's mouth. A loudspeaker (Boston A40V), placed on top of the monitor behind the middle panel, reproduced the audio stimulus. During the experiment the room lights were extinguished; the only light available was that generated by the video image.

E. Procedure

Infants were seated in the infant seat which was firmly clamped to a 76-cm-high table. The child was secured in the infant seat with Velcro straps. When seated, infants faced the three-sided cloth-covered theater, the front panel of which displayed the video screen. Before the experiment, infants were held by the parent in the room for a few minutes to adjust to the room. Then, the mother handed the infant to an assistant, left the room, and closed the door to the soundproof room. The stimulus tape was started to allow the infant to adjust to the voice of the talker while the assistant secured the infant in the infant seat. Once the infant was secured, the assistant walked behind the infant and pressed a button which initiated the 5-min recording period. During the 5-min session, the assistant did not talk to or touch the infant. At the completion of the 5-min recording period, the infant was removed from the seat.

F. Scoring and acoustic analysis of infants' vocalizations

The analysis of vocal responses had three components: the selection of vocalizations, perceptual analysis of the vocalizations, and instrumental analysis of the vocalizations.

All vocalizations that occurred in the silent intervals between the adult's vocalizations (3.37 s, on average) were isolated for further analysis. Adult utterances (and all infant utterances that overlapped with those of the adult) were eliminated for two reasons: (a) when analyzing the vocalizations perceptually, hearing the voice of the adult would potentially bias the observer toward the category of sounds produced by the model, and (b) when analyzing signals instrumentally there is no way of separating the utterances produced by the adult from those produced by the infant. The elimination of the adult's vowels was done instrumentally by sending the audio signal recorded from the test room to two devices, the video recorder and the audio cassette recorder. The audio cassette recorder received all audio information recorded in the room, that is, both the model's and the infant's vocalizations. However, the video recording, which was used by the transcriber, received only the infants' utterances. The models' utterances were blanked out of the audio portion of the video recording by a specially constructed hardware/software interface between the computer and the video recorder that shut down the audio recording 50 ms before until 50 ms after the adult's utterance. The result is that no adult vocalizations occurred on the scoring tape, and therefore no possibility of scorer bias.

"Criterion" utterances were selected by a trained phonetician who watched the video and listened to the audio track that contained only the infants' utterances. The scorer was thus completely unaware of the stimulus presented to the infant. The scorer selected utterances that contained "vowel-like" sounds on the basis of the articulatory and perceptual characteristics typical of vowels. The operational definition of vowel-like was a continuous, voiced sound produced with normal laryngeal vibration in the absence of aspiration or frication. Utterances had to be produced on an exhalatory breath with a visibly open mouth, be relatively steady state, and have an audible voice pitch. Utterances that occurred while the infant's hands were in the mouth were eliminated. The application of these operational criteria ensured that a range of nonspeech vocalizations were eliminated such as crying, laughing, sneezing, squealing, smacking, coughing, gurgling, grunting, hiccoughing, and burping. What remained were the sounds typically referred to in the literature as infant "cooing."

Of the 72 infants tested, 63 produced at least one vocal utterance. Of these 63 infants, 45 produced at least one criterion utterance qualifying as vowel-like using the operational definition. For the group of 45 infants, a total of 224 utterances met the criteria (M=4.98 per infant, range: 1–27).

These 224 utterances were scored both perceptually and instrumentally. The perceptual analyses were done by having an individual trained in the phonetic transcription of infants' vocalizations listen to each of the utterances and code it as one of the eight vowels: /i, ɪ, ɛ, æ, a, ʌ, ʊ, u/. Although young infants do not produce perfect instances of these vow-

els, a person trained in coding infants' vocalizations can judge which vowel category an utterance most closely resembles. For scoring purposes, the 224 utterances were randomly ordered on a test videotape. When the scorer transcribed the utterances from this tape, she could replay utterances as often as she wanted before making a judgment. Once transcribed, the vowels were statistically analyzed by grouping the eight phonetic categories into one of three groups: /a/-like vowels (/a/, /æ/, and /ʌ/), /i/-like vowels (/i/, /ɪ/, and /ɛ/), and /u/-like vowels (/ʊ/ and /u/). To assess reliability, a second phonetically trained individual rescored the entire corpus of utterances. Transcriptional agreement for coding utterances into the three categories was 86%, calculated as the number of disagreements divided by the sum of the agreements+disagreements.

Instrumental analysis was conducted by an individual who was blind to the transcriptional classification of each utterance. This person was trained in speech acoustics and in the use of specialized spectrographic equipment. The individual had prior experience analyzing infants' vocalizations, which are particularly difficult to analyze because of infants' high fundamental frequency and the attendant low density of harmonics in the source spectrum. Infants' vocalizations were lowpass filtered at 10 kHz and digitized at a 20-kHz sampling rate. The infant utterances were analyzed using Kay Elemetrics microcomputer-based equipment (CSL, version 4.0). Formant values were obtained through the corroboration of three analyses, a narrowband spectrogram (114 Hz), a fast Fourier transform (256 points, preemphasis, Blackman window weighting), and LPC frequency response (10-ms frame length, filter order=12).

Each utterance was sampled at five locations: onset, at the 1/4, 1/2, and 3/4 points, and offset. Five acoustic parameters were measured at each location: the first formant ($F1$), the second formant ($F2$), the compact-diffuse (CD) feature, the grave-acute (GA) feature, and the fundamental frequency ($F0$). Total duration (DUR) was also measured. The CD and GA features distinguish /a/, /i/, and /u/ vowels, and specify the relationship between the first two formant frequencies. CD measures the relative spacing between the first two formant frequencies and was calculated by subtracting $F1$ from $F2$. GA measures the extent to which the average of the first two formant frequencies is low or high in frequency, $(F1+F2)/2$. According to feature theory (Jakobson *et al.*, 1969), the vowel /a/ is compact and grave, while the vowel /i/ is diffuse and acute. These acoustic features of speech are not frequency specific, and using them makes it possible to compare the vocalizations of adults and infants. Interscorer agreement on the spectrographic measures was estimated by having an additional instrumentally trained person repeat 20% of the measures drawn randomly across vowels, subjects, locations, and acoustic parameters ($F1$, etc.). The data revealed that the measurement error was within 5% of the original value, which is considered excellent for infant vocalizations (Kent, 1992).

II. RESULTS

Infants' vocalizations were analyzed both perceptually and instrumentally. Perceptual analysis was accomplished by phonetically trained listeners and instrumental analysis by spectrographic measurement. The data bear on three questions: (a) Where do infant vowel productions fall within adult vowel space? (b) Are there developmental changes in vowel production between 12- and 20-weeks of age?, and (c) Do young infants exhibit vocal imitation?

A. Infant vowel space: Relating infants' to adults' vocalizations

Infants produced 224 utterances meeting the criteria for vowel-like vocalizations. The number of utterances as a function of age (12-, 16-, and 20-weeks old) was, respectively, 79, 83, and 62. The frequency of different utterance types (/a/-like, /i/-like, and /u/-like vowels) in the corpus was respectively, 105, 37, and 82.

The values of $F1$, $F2$, GA, CD, and $F0$ were obtained by instrumental measurement for each utterance. Statistical tests indicated that there were no significant differences in the parameter values depending on whether the middle three locations within an utterance (1/4, 1/2, or 3/4 points) or all five locations were considered. Therefore the average of five locations was used to specify each parameter ($F1$, $F2$, etc.) in the table, figures, and statistical analyses which follow.

A total of 5824 acoustic measurements (26 measurements per utterance×224 utterances) were attempted across the corpus. In the case of 34 utterances, acoustic measurement was not possible due either to the presence of noise (Velcro noise from the strap that secured the infant in the seat), or the fact that the infant's fundamental frequency was high and energy was present at all harmonics obscuring the formant frequencies. Although these factors did not prevent phonetic transcription (because the transcriber could perceptually segregate utterances from noise and identify vowel quality for utterances with high fundamental frequencies), these factors prevented accurate spectrographic measurement. These 34 utterances were therefore not analyzed acoustically. Descriptive statistics for each of the six acoustic dimensions ($F1$, $F2$, CD, GA, $F0$, and DUR) on the resulting corpus of 190 utterances are provided in Table I. Table I lists the mean, standard deviation, minimum, and maximum for each utterance type at each age.

Figure 1 displays the entire corpus of infant vocalizations. Each utterance (as transcribed by vowel category) is cast in an $F1/F2$ coordinate plot. The values of $F1$ in the infant corpus range from 487 to 1645 Hz; $F2$ from 1523 to 4120 Hz. The durations range from 178 to 2195 ms. Both the formant values and the durations are consistent with those reported in the one other study in which young infants' vowel-like vocalizations were instrumentally measured (Kent and Murray, 1982). As shown, the utterances transcribed as members of each particular vowel category are clustered. Moreover, the three vowel clusters are positioned in a way predicted by the acoustic measurement of adults' vowels; $F1$ and $F2$ values for the three categories are in the appropriate relationship to one another (Peterson and Barney, 1952).

Figure 2 plots our corpus of infant vowels within the vowel space reported in the classic study of Peterson and Barney (1952). In the Peterson and Barney study, the vowels

TABLE I. Means, standard deviations, and ranges for the acoustic measures of infants' vowels as a function of age.

Age	Utterance	Mean	SD	Minimum	Maximum
			*F*1 (Hz)		
12	/a/	933.9	205.8	598.4	1645.0
	/i/	781.6	106.8	646.8	950.0
	/u/	731.7	113.7	537.0	903.0
16	/a/	1044.4	135.5	729.4	1516.8
	/i/	739.1	3.8	736.4	741.8
	/u/	757.0	97.9	585.8	1057.5
20	/a/	945.6	156.5	706.0	1169.0
	/i/	778.3	111.7	596.6	998.6
	/u/	675.1	89.9	486.6	844.8
			*F*2 (Hz)		
12	/a/	2606.4	474.2	1832.8	3637.6
	/i/	3121.2	337.5	2700.6	3863.0
	/u/	2198.7	502.2	1608.4	4119.5
16	/a/	2499.4	372.9	1810.3	3613.2
	/i/	2887.3	660.4	1523.0	2457.0
	/u/	2156.2	209.7	1683.4	2530.0
20	/a/	2393.7	309.6	1863.0	2940.8
	/i/	2947.3	473.1	2085.5	3879.0
	/u/	2335.1	141.6	2131.8	2612.0
			Compact-Diffuse (Hz)		
12	/a/	1685.7	451.1	1013.8	2684.2
	/i/	2334.6	332.1	1750.6	3010.8
	/u/	1464.4	507.3	901.4	3282.8
16	/a/	1452.4	372.2	773.8	2767.8
	/i/	2139.7	667.4	771.3	1715.2
	/u/	1398.5	203.2	1023.0	1769.0
20	/a/	1472.6	201.6	1130.2	1771.9
	/i/	2171.5	503.7	1142.8	3115.4
	/u/	1660.1	139.2	1525.0	2024.2
			Grave-Acute (Hz)		
12	/a/	1763.6	275.7	1215.6	2295.5
	/i/	1953.9	186.3	1783.3	2357.6
	/u/	1466.5	261.1	1120.8	2478.1
16	/a/	1772.0	209.2	1363.4	2229.3
	/i/	1817.4	326.7	1137.4	1599.4
	/u/	1457.0	128.4	1171.9	1647.8
20	/a/	1657.5	225.4	1297.9	2054.9
	/i/	1864.0	234.7	1406.3	2321.3
	/u/	1505.1	96.0	1315.9	1653.8
			*F*0 (Hz)		
12	/a/	316.2	39.4	264.4	433.0
	/i/	336.5	41.0	253.3	392.0
	/u/	323.6	39.4	256.8	406.8
16	/a/	318.9	31.9	265.0	422.3
	/i/	311.2	20.6	296.6	325.8
	/u/	341.0	28.0	260.3	395.3
20	/a/	296.4	19.9	272.3	337.4
	/i/	321.5	45.7	245.2	404.5
	/u/	324.3	24.5	294.3	365.4
			Duration (s)		
12	/a/	0.526	0.261	0.244	1.347
	/i/	0.438	0.114	0.326	0.630
	/u/	0.544	0.333	0.257	2.195
16	/a/	0.730	0.376	0.254	1.822
	/i/	0.576	0.302	0.363	0.790
	/u/	0.679	0.332	0.255	1.366
20	/a/	0.683	0.326	0.381	1.422
	/i/	0.479	0.241	0.178	1.186
	/u/	0.561	0.307	0.185	1.073

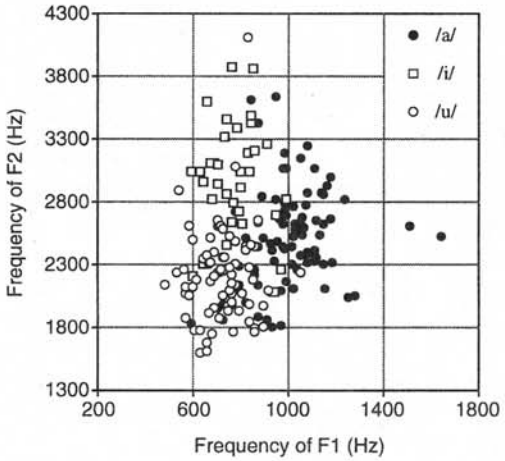

FIG. 1. The corpus of infant utterances plotted in an *F*1 versus *F*2 coordinate vowel space. Infants' utterances are coded by vowel category (/a/-like, /i/-like, or /u/-like) as determined by phonetic transcription.

of 76 speakers, including men, women, and child speakers of American English were measured. The closed curves shown in Fig. 2 are Peterson and Barney's, drawn by their visual inspection to encompass 90% of the utterances in each vowel category. Superimposed on the graph is a closed curve enclosing approximately 98% of the utterances from the infant corpus obtained in the present study. As shown, infants' vowels overlap with certain adult vowel categories (particularly /ɛ/ and /æ/) but extend the vowel space considerably beyond that used by adult and child speakers. This is as expected; infants' vocal tracts are smaller and this results in higher resonant frequencies and correspondingly higher formant frequencies. While the infant vowel space is more restricted than that of the adult, illustrating that the infant's vocal tract is not anatomically capable of producing the full

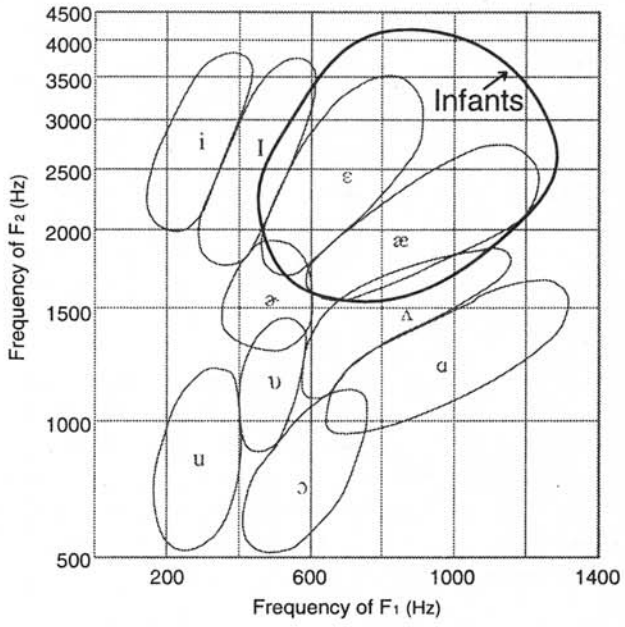

FIG. 2. The "vowel space" of 12-, 16-, and 20-week-old infants in relation to the plot published by Peterson and Barney (1952) that was based on vowel productions of 76 men, women, and children.

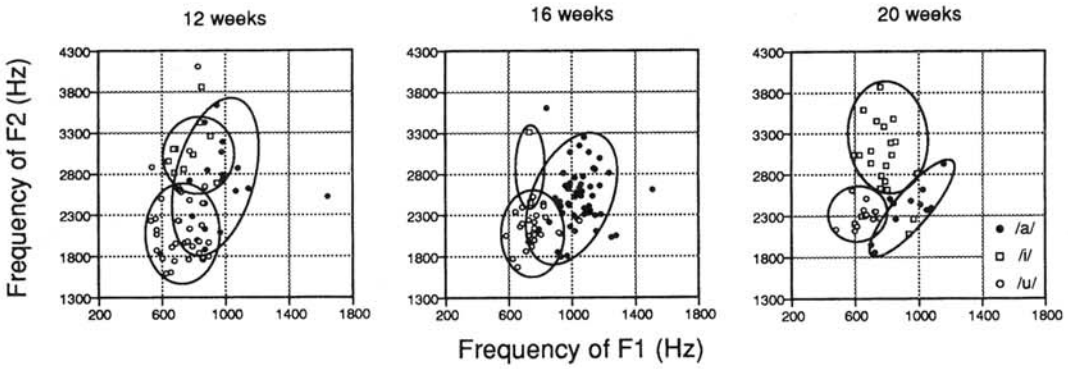

FIG. 3. The /a/-like, /i/-like, and /u/-like vowels produced by 12-, 16-, and 20-week-old infants cast in $F1$ versus $F2$ coordinate plots. The closed curves were drawn by visual inspection to enclose 90% or more of the infants' utterances. Infants' vowel categories show greater separation in vowel space as a function of age.

range of formant frequency variation seen in adult and child speakers, they nonetheless produce vowel-like sounds that show substantial acoustic variation.

B. Developmental changes in vocalizations between 12 and 20 weeks of age

Figure 3 displays the vowels of 12-, 16-, and 20-week-old infants in $F1/F2$ coordinate spaces. In each graph, infants' vowel utterances are coded according to the transcription provided by the phonetically trained listener. The closed curves, drawn by visual inspection of the graphs, enclose 90% or more of the utterances in each category.

It is clear from Fig. 3 that utterances coded as a particular vowel form a cluster in acoustic space. This is the case even for the youngest infants, the 12-week-olds. For example, vowels with the highest $F2$ values and relatively low $F1$ values are coded as /i/, while those with lowest $F1$ and $F2$ values are coded as /u/. The relationship between the acoustic properties and the transcription observed for infants' vowels is similar to the relationship between acoustic properties and transcription that exists in the categorization of adults' vowels (Peterson and Barney, 1952).

Figure 3 also reveals that the areas of vowel space occupied by infants' /a/-, /i/-, and /u/-like vowels become progressively more *separated* between 12 and 20 weeks of age, due to a tighter clustering of the vowels in each category over time. This developmental shift could be due to anatomical changes that stabilize infants' articulatory movements over time. This would be compatible with the fact that infants' vocal tracts are rapidly changing during this period (Sasaki *et al.*, 1977). On the other hand, it is intriguing to consider the possibility that the increasingly tighter clustering seen for categories of infants' vowels could be due, at least in part, to vocal learning. Perhaps infants are listening to the vowels produced by adult speakers of the language and are striving to produce vowels themselves that perceptually resemble those they hear adults produce. This latter point hinges on infants' abilities to match, with their own vocalizations, the vocalizations they hear another person produce (discussed below).

It is also of interest to examine developmental change in infants' vocalizations using the featural measures. GA and CD measurements were taken for each of the vowels at each of the five locations. The GA and CD features distinguish adults' vowels, especially the vowels /a/, /i/, and /u/ (see Syrdal and Gopal, 1986). Figure 4 displays the featural measures of the utterances of 12-, 16-, and 20-week-old infants in compact-diffuse/grave-acute coordinate plots. The closed curves encircling the utterances of a particular type were drawn by visual inspection to encompass 90% or more of the

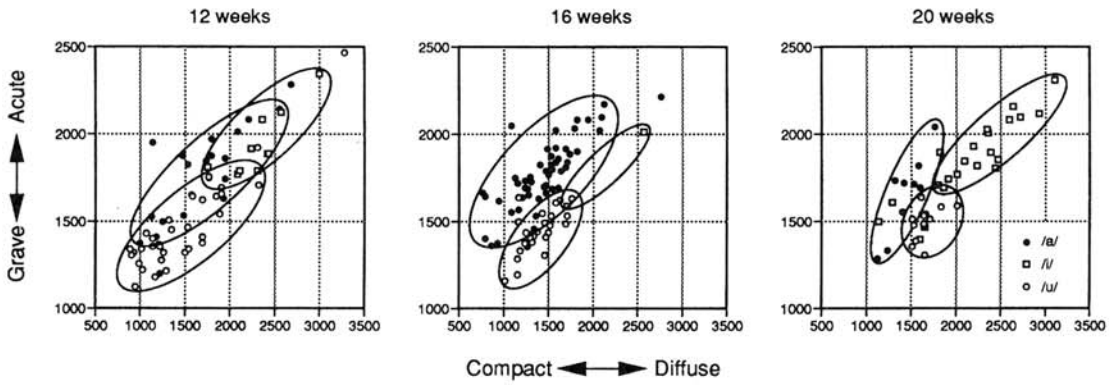

FIG. 4. The /a/-like, /i/-like, and /u/-like vowels produced by 12-, 16-, and 20-week-old infants cast in compact/diffuse versus grave/acute coordinate plots. The closed curves were drawn by visual inspection to enclose 90% or more of the infants' utterances. Infants' vowel categories show greater separation in vowel space as a function of age.

utterances in each category. These plots suggest that, at the earliest age tested, 12 weeks, the /a/, /i/, and /u/ utterances are differentiated by the GA and CD acoustic features. Examination of the space occupied by the three vowels across age, however, suggests that infants' vowel categories become much more separated over this 8-week period.

C. Statistical analysis of the acoustic measurements

The plots in Figs. 3 and 4 suggest that the acoustic measures differentiate infants' vowel categories (as defined by the transcriber). Statistical analysis of each acoustic variable was undertaken to verify which measures were statistically reliable. Using a 3 (age: 12, 16, 20)×3 (vowel category: /a/, /i/, /u/) analysis of variance (ANOVA), the main effects and interactions were examined for each of the six acoustic measurements ($F1$, $F2$, CD, GA, $F0$, and DUR). When appropriate, follow-up tests were conducted (simple effects and Tukey-HSD post hoc tests).

Analysis of the $F1$ measurements revealed a significant main effect of vowel category, $F\ (2,181)=66.69$, $p<0.001$, and a main effect of age, $F\ (2,181)=5.75$, $p<0.005$. The interaction between age and vowel category was not significant, $p>0.20$. Follow-up tests revealed that at 12 and 16 weeks, the $F1$ values for /a/ were significantly higher than for /i/ or /u/; the values of /i/ and /u/ did not differ. By 20 weeks of age, tests revealed that the $F1$ values for /a/, /i/, and /u/ all differed significantly. Thus, analysis of the $F1$ data reveals that at each age significant differences in the predicted direction exist among the /a/, /i/, and /u/ utterance types; moreover, the data show that the $F1$ values of /a/ are separated from /i/ and /u/ before the latter two are differentiated.

Analysis of the $F2$ measurements revealed a significant effect of vowel category, $F(2,181)=40.75$, $p<0.001$. Neither the effect of age, $p>0.40$, nor the interaction between age and vowel category, $p>0.40$, was significant. Follow-up tests revealed that the $F2$ values for /i/ are significantly higher than those for /a/ and /u/ at all ages. The pattern of $F2$ values shown by infants thus conforms to that shown by adults.

Analysis of the CD measurements revealed a significant effect of vowel category, $F(2,181)=32.64$, $p<0.01$. Neither age, $p>0.08$, nor the interaction between age and vowel category, $p>0.20$, was significant. Follow-up tests revealed that the vowel /i/ is significantly more diffuse than /a/ and /u/, as expected, for all three ages.

Analysis of the GA measurements revealed a significant effect of vowel category, $F(2,181)=52.03$, $p<0.001$. Neither the effect of age nor the interaction between age and vowel category was significant, $p>0.40$ in both cases. Follow-up tests show that /u/ is significantly more grave than either the /i/ or /a/ vowels. Thus, the patterning of the three vowels on the GA dimension conforms to the pattern shown by adults.

Analysis of the $F0$ measurements revealed a significant effect of vowel category, $F(2,181)=4.61$, $p<0.05$. Neither the effect of age nor the interaction, between age and vowel category was significant, $p>0.10$ in both cases. Follow-up tests revealed that the significant effect was attributable to the fact that the 16-week-olds produced their /u/ vowels with a higher $F0$ than either their /i/ or /a/ vowels. No other significant differences in $F0$ were observed.

The main effects of vowel category and age were also examined for the duration (DUR) measurements. We had not predicted any specific variations in duration as a function of either vowel category or age. The analysis revealed that the effect of vowel category was not significant, $p>0.10$, but that the effect of age was significant, $F(2,181)=4.23$, $p<0.02$. The interaction was not significant. Follow-up tests showed that 16-week-olds' utterances were significantly longer than those of either the 12- or the 20-week-olds.

Taken as a whole the acoustic measurements show significant effects of both vowel category and age on infants' vocalizations. Infants' vowels, even though they occupy a smaller area within the vowel space when compared to adults' vowels, differ perceptually (as shown by phonetic transcription). Moreover, the perceptual differences in infants' vowels correlate with acoustic differences that are consistent with the acoustic dimensions that differentiate adults' vowels (Peterson and Barney, 1952; Syrdal and Gopal, 1986).

D. Vocal imitation

The analyses thus far have demonstrated that young infants produce vowels that phonetically trained listeners can reliably code as /a/-like, /i/-like, and /u/-like, and that these categories vary acoustically, both in their formant values and in acoustic features calculated from the formant values. The fact that infants are capable of producing utterances perceived as /a/-like, /i/-like, and /u/-like by adult listeners allows us to pose the next question: Do infants systematically vary the utterances they produce as a function of the stimulus they hear? In other words, Is there evidence that infants are attempting to imitate the adult model?

If infants are capable of vocal imitation their vocalizations should vary as a function of the three stimulus conditions. The hypothesis of vocal imitation predicts that infants should produce more /a/-like utterances in response to the stimulus /a/ than to the /i/ or /u/ stimulus; similarly, they should produce more /i/-like utterances in response to the stimulus /i/ than to the /a/ or /u/; and finally, they should produce more /u/-like utterances in response to the stimulus /u/ than to the /a/ or /i/. Data regarding imitation will be presented at two levels of analysis, both at an utterance level and at the level of individual subjects.

The utterance-level analysis examines the entire corpus of 224 utterances. Figure 5 displays the corpus of 224 utterances in a stimulus-response matrix that provides the number of each infant utterance type (/a/-like, /i/-like, or /u/-like) as a function of stimulus condition (/a/, /i/, or /u/). Recall that infants were assigned randomly to stimulus groups, and that each infant was exposed to a video of the same female producing a vowel at the same rate for the same length of time. The only thing that varied was the particular vowel she produced. If the stimulus had no effect on infants' productions, the cells in a given row should not vary. Alternatively, if the stimulus affects infants' responses and infants are attempting to match the stimulus, then the largest cell frequencies will

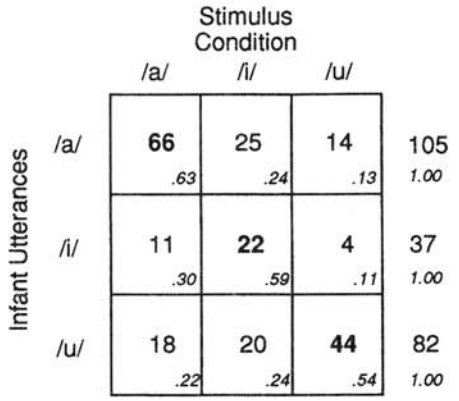

FIG. 5. Stimulus × response matrix for the entire corpus of 224 utterances showing number of infant utterances (/a/-like, /i/-like, or /u/-like) occurring in response to the stimulus (adult presentation of /a/, /i/, or /u/). Higher numbers along the diagonal support the hypothesis of vocal imitation.

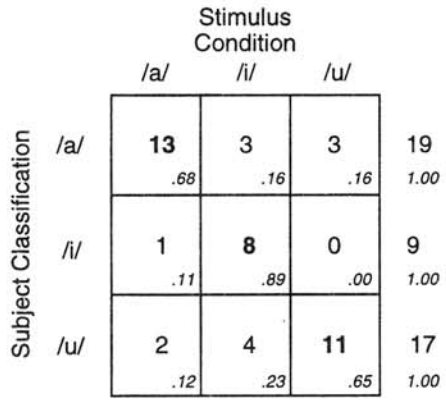

FIG. 7. Stimulus × subject classification matrix for 45 infants. Cell entries are the number of infants classified as /a/ infants, /i/ infants, or /u/ infants (based on their vocalizations) as a function of stimulus condition (adult presentation of /a/, /i/, or /u/). Higher numbers along the diagonal support the hypothesis of vocal imitation.

occur on the diagonal. The frequency of /a/ utterances will be at its maximum when the stimulus is /a/; the frequency of /i/ utterances will be maximum when the stimulus is /i/, and the frequency of /u/ will be maximum when the stimulus is /u/.

Visual inspection of the cell frequencies in the stimulus-response matrix of Fig. 5 suggests that the vowel stimulus strongly affected the type of vowel infants produced in response. In all three rows, the cell with the highest frequency falls on the diagonal. The top row shows that the frequency of /a/ utterances was systematically influenced by the stimulus that infants heard. Of the 105 /a/ utterances produced by infants in the experiment, 66 (62.9%) occurred in response to the stimulus /a/, 25 (23.8%) occurred in response to the stimulus /i/, and 14 (13.3%) occurred in response to the stimulus /u/. A similar pattern emerges in the case of /i/ utterances. Of the 37 /i/ utterances produced by infants, 22 (59.4%) occurred in response to the stimulus /i/; 11 (29.7%) occurred in response to the stimulus /a/, and 4 (10.8%) occurred in response to /u/. The /u/ utterances also showed a similar profile. Of the 82 /u/ utterances produced by infants, 44 (53.7%) occurred in response to the stimulus /u/, 18 (22.0%) occurred in response to /a/, and 20 (24.4%) occurred in response to /i/.

Figure 6 displays the stimulus-response matrices for each age individually. Examination of the nine rows in these tables (3 ages×3 response categories) shows that eight of the nine are in line with the prediction of vocal imitation.

The utterance-level data just presented are informative because they show the distribution of all 224 utterances. However, infants' utterances entered into the table cannot be considered independent of one another. Therefore, a second analysis was undertaken at the subject level. In this analysis, infants were categorized as ''/a/ infants,'' ''/i/ infants,'' or ''/u/ infants'' depending on the utterance type they most frequently produced. For example, an infant who produced 12 criterion utterances during the course of the experiment, 7 coded as /u/-like, 3 as /a/-like, and 1 as /i/-like, would be classified as a ''/u/ infant,'' and so on. In classifying infants, only utterances transcribed identically by the two transcribers were considered. This was done to ensure that only infants' clearest utterances would be used in determining their classification as an /a/, /i/, or /u/ infant. The subject-level classification achieves statistical independence because each infant is listed once and only once in the matrix.

Figure 7 displays the subject-classification×stimulus-condition matrix for all 45 infants who produced criterion

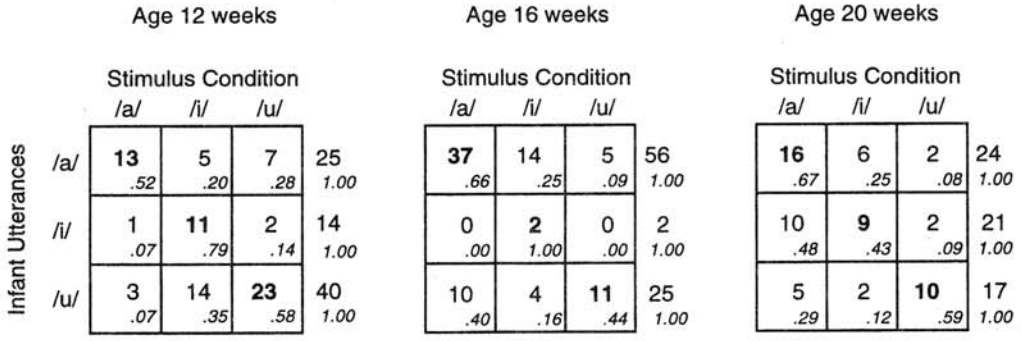

FIG. 6. Stimulus × response matrices for each individual age. Cell entries are the number of infant utterances (/a/-like, /i/-like, or /u/-like) that occurred in response to the stimulus (adult presentation of /a/, /i/, or /u/). Higher numbers on the diagonal support the hypothesis of vocal imitation.

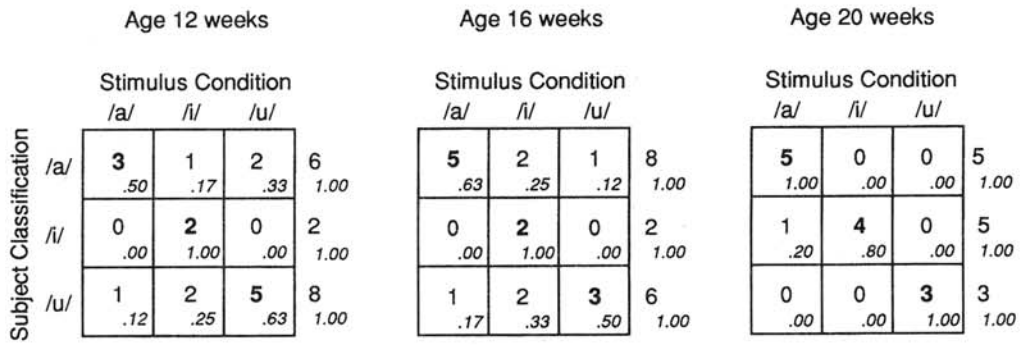

FIG. 8. Stimulus × subject classification matrices for each individual age. Cell entries are the number of infants classified as /a/ infants, /i/ infants, or /u/ infants (based on their vocalizations) as a function of stimulus condition (adult presentation of /a/, /i/, or /u/). Higher numbers along the diagonal support the hypothesis of vocal imitation.

utterances during the experiment. These data demonstrate that the vowel infants heard systematically affected their classification. In all three rows, the cell with the highest frequency falls on the diagonal, supporting the hypothesis of vocal imitation. A chi-square test of the 3×3 contingency table is significant, $\chi^2(4, N=45)=30.97$, $p<0.0001$.[1] Each row of Fig. 7 can also be considered individually. The top row shows that of the 19 infants classified as /a/ infants, 13 (68.4%) had heard the /a/ vowel during the experiment, 3 (15.8%) had heard /i/, and 3 (15.8%) had heard /u/, $\chi^2(2, N=19)=10.53$, $p<0.005$. A similar pattern emerges in the case of /i/ infants. Of the 9 infants classified as /i/ infants, 8 (88.9%) had heard the /i/ vowel during the experiment, and only 1 (11%) had heard the /a/ vowel, $p<0.025$ by the binomial test. A similar pattern is also obtained in the case of /u/ infants. Of the 17 infants classified as /u/ infants, 11 (64.7%) had heard the vowel /u/ during the experiment, 2 (11.8%) had heard /a/, and 4 (23.5%) had heard /i/, $\chi^2(2, N=17)=7.88$, $p<0.025$.

Figure 8 displays the subject-level matrices for each age considered individually. Examination of the nine rows (3 ages×3 response categories) shows that all are in line with the prediction of vocal imitation. Chi-square analyses showed that the matrix for the 20-week-olds reached significance, $\chi^2(4, N=13)=21.67$, $p<0.001$. The 12- and 16-week-old matrices did not reach significance when considered individually; however, the data were significant with the larger N provided by collapsing the two younger age groups (12- and 16-week-olds) together, $\chi^2(4, N=32)=15.54$, $p<0.01$. The subject-level analyses support the hypothesis of vocal imitation in infants under 20 weeks of age.

III. DISCUSSION

In this experiment infants' vocalizations in response to speech were examined. Infants at three ages, 12, 16, and 20 weeks of age, were recorded while they saw and heard a woman producing one of three vowels, /a/, /i/, or /u/. Infants' vocalizations were analyzed perceptually by having them phonetically transcribed and analyzed instrumentally using computerized spectrographic techniques. The discussion addresses three points: (a) infants' vowel space and the relationship between infants' vowels and adults' vowels, (b) developmental changes in infants' vowel productions, and (c) vocal learning, imitation, and the intermodal nature of speech in infancy.

A. Infants' vowel space

Acoustic analyses showed that although approximately 50% of infants' vowels overlapped the area of vowel space used by children and adults, half extended considerably beyond the traditional vowel space plotted by Peterson and Barney (1952). Infants produced vowels with higher $F1$ and $F2$ values than observed in the vowels of adult and child speakers of English. This was not unexpected given the smaller vocal tracts of infants and the resulting increase in the formant frequencies. From a descriptive standpoint, it is helpful to map the area used by infants so that future work may examine how it changes with age. From a more theoretical viewpoint, the data raise the issue of how listeners (including infants) perceive "constancy" across the vowels produced by different people.

Consider the category of /u/ vowels. The /u/-like vowels of infants have very different formant frequencies than the /u/ vowels spoken by men, women, and child speakers of English. How do adults (or, for that matter, infants) hear all these different /u/s as similar? Experiments have established that 6-month-old infants hear the similarity among vowels spoken by male, female, and child talkers (Kuhl, 1979, 1983, 1985). More recent experiments suggest that younger infants, 2-month olds (Marean *et al.*, 1992) and newborns (Walton *et al.*, 1991), perceive vowel constancy as well. The results of the present experiments show that adults can also reliably classify infants' vowel-like utterances into phonetic categories. As underscored by these results and by others, the identity of a vowel is not determined by its absolute location in vowel space. Various authors have suggested that vowel constancy is achieved by relational information that takes into account the fundamental frequency of the vowel and the relationships among formants contained in the vowel (Miller, 1989; Nearey, 1989; Syrdal and Gopal, 1986). There is as yet no full account of the underlying mechanism in humans, but the data suggest vowel constancy is a basic ability infants bring to the task of acquiring language.

Vowel constancy is critical to infants' speech and language development. Our data demonstrate that many infant

vowels are outside the frequency range produced by adult speakers. If infants were unable to recognize the similarity among the vowels produced by different people, they would be unable to relate their own vowel productions to the vowels of adult speakers. It would not be possible for infants to learn the "mother tongue" without relating their own vowel-like sounds to those produced by their language tutors. The dual findings from infant vowel categorization studies (Kuhl, 1979, 1983, 1985; Marean *et al.*, 1992; Walton *et al.*, 1991) and the present data on imitation provide converging support for the inference that young infants have in place a perceptual-motor system that allows them to use vowels spoken by adults to guide the production of their own vowels.

B. Developmental changes in vowel production: Extending the native language magnet model to speech production

The acoustic analysis of infants' vowels demonstrated developmental change. From 12 weeks of age to 20 weeks of age, infants' vowel categories became increasingly differentiated. This was true whether infants' vowels were examined in a two-formant coordinate vowel space or in a compact-diffuse, grave-acute vowel space. Infant vowel categories appear to be more tightly clustered at 16 weeks than at 12 weeks, and more tightly clustered at 20 weeks than at 16 weeks. What causes the increased separation of vowel categories over this relatively short (8-week) period?

Anatomical changes could play a role. However, what kind of anatomical change could be responsible for the precise shift observed in infants' vocalizations? There was no general shift observed in the frequency of $F1$ or $F2$ in infants' vocalizations. Rather, tighter clustering in three areas of the vowel space which resulted in greater separation between the vowel categories was observed. Anatomical factors alone would be hard pressed to account for these specific changes.

We suggest a second account for the observed clustering of vowels in acoustic space. We propose that infants' perceptual representations of vowels, stored in memory, serve as "targets" that infants try to match when producing speech themselves. According to Kuhl's native language magnet (NLM) model, infants store representations of the vowels of their native language (Kuhl, 1992, 1993a, b, 1994; Kuhl and Meltzoff, in press). These stored representations derive from linguistic input and alter speech perception at a very early age. The hypothesis being put forward here extends the NLM model to developmental change in speech production.

How might this work? Studies of infant speech perception demonstrate that linguistic experience alters speech perception at a very early age (Grieser and Kuhl, 1989; Kuhl, 1991; Kuhl *et al.*, 1992). In Kuhl *et al.* (1992), 6-month-old infants in the United States and Sweden were tested with English and Swedish vowel "prototypes," exceptionally good instances of the two vowel categories. The results showed that both groups of infants treated the native-language sound in a unique way: They demonstrated greater perceptual generalization (or clustering) around the native-language prototype than around the foreign-language prototype, even though psychophysical distance was equated. Kuhl (1991) termed the greater clustering around prototypes a *perceptual magnet effect*. The effect indicates that vowel space is warped by linguistic experience. Psychological distance in the vicinity of the category prototype is shrunk when compared to a nonprototype (Iverson and Kuhl, 1995; Iverson and Kuhl, 1996; Kuhl, 1991; Kuhl and Iverson, 1995). Kuhl (1992, 1993a, b, 1994) suggested that the magnet effect is attributable to memory representations of speech sounds.

Even if exposure to ambient language results in representations for speech sounds, infants' developmental changes in speech production still require some sort of link between the stored representations and motor output. We think we now have such evidence. The current study demonstrated that a total of 15 min of exposure (5-min exposure to a specific vowel for each of 3 days) was sufficient to influence vocalizations in infants under 20 weeks of age. If 15 min of laboratory exposure to a specific vowel is sufficient to influence infants' production of that vowel, then listening to the ambient language for many weeks could plausibly provide sufficient exposure to induce the kind of change seen in infants between 12 and 20 weeks of age.[2]

It is known from the results of babbling studies in different cultures that very long-term exposure to speech influences infant speech production (de Boysson-Bardies *et al.*, 1989; de Boysson-Bardies *et al.*, 1984). Two-year-olds from different cultures clearly sound different. The current experimental study demonstrates that controlled laboratory exposure influences the vocal productions of infants under 20-weeks of age. Thus, the potential for vocal learning in the early months of life exists and may be capitalized on in more natural environmental interactions.

We thus seek to unify recent findings in speech perception and production by suggesting that representations of speech may underlie both. The hypothesis is that the tighter clustering observed in the present study of infant vowel *production* and previous results showing tighter clustering in infant vowel *perception* as a function of linguistic exposure are both attributable to a common cause—the formation of memory representations that derive initially from perception of the ambient input and then act as guides for motor output (Kuhl and Meltzoff, in press).

C. Vocal imitation, vocal learning, and the intermodal nature of speech

Vocal imitation for infants between 12 and 20 weeks was demonstrated. When listening to an adult speaker produce vowels, infants responded with vocalizations that perceptually matched the vowels presented to them. Infants produced more /a/-like utterances when listening to /a/ than they did when listening to /i/ or /u/; similarly, infants produced more /i/-like vowels when listening to /i/ than when listening to /a/ or /u/; finally, they produced more /u/-like vowels when listening to /u/ than when listening to either /a/ or /i/. Subject-level analyses that classified infants as /a/, /i/, or /u/ infants according the their predominant vocalizations provided direct statistical support for the vocal imitation hypothesis. Instrumental analysis revealed that infants' /a/-like, /i/-like, and /u/-like vowels exhibit the same featural relationships as exist in the /a/, /i/, and /u/ vowels produced

by adults. The combined transcriptional and instrumental results provide strong support for the hypothesis that infants between 12 and 20 weeks are capable of vocal imitation.

Was imitation guided by the facial movements infants saw or by the sounds they heard? As yet, studies have not been conducted isolating the visual component of the experiment (the face without sound) from the auditory component (the sound without the face). However, one study is relevant. When the face used in this study was presented with nonspeech tones sharing one spectral feature of the vowels (but not identifiable as speech), infants did *not* produce speech sounds (Kuhl and Meltzoff, 1982, 1988). This suggests that hearing a speech sound, not merely seeing the moving face, plays a role in compelling infants to produce speech themselves (see also Legerstee, 1990).

How is it that by 12 weeks of age infants are capable of moving their articulators in a way that is appropriate to achieve a specific auditory target? We do not know whether there is an initial tendency for human speech to drive infants' vocal productions analogous to the tendency for visually perceived body movements to drive corresponding body acts, as manifest in newborn gestural imitation (Meltzoff, 1990; Meltzoff and Moore, 1977, 1983a, 1992, 1994). Even if so, speech development would rapidly be influenced by experience, especially that gained by infants' own cooing and sound play. Infant cooing begins at about 4 weeks of age and would allow infants to discover that articulatory movements of a particular type have very specific auditory consequences. This would result in the development of an auditory–articulatory "map" relating self-produced auditory events to the motor movements that caused them. These experiences would, in turn, lead to the development or refinement of any perceptual-motor linkages that were present initially.

Presumably, infants' successive approximations of vowels would become more accurate as infants relate the acoustic consequences of their own articulatory acts to the acoustic consequences of the articulatory acts of others (representations stored in memory). We argue that infants make progress by comparing the two, eventually converging on vowels that match their memory representations. Infants in particular cultures would thus begin to produce utterances that resemble ambient language input.

This account implies that infants not only have to be able to hear the sounds produced by others, but that they need to hear the results of their own attempts to speak in order to learn to speak (Locke, 1993). Without the auditory feedback from their own articulatory movements, they would not close the perceptual-motor loop and fill in the auditory–articulatory map. Both hearing the sound patterns of ambient language (auditory exteroception) and the ability to hear one's own attempts at speech (auditory proprioception) are critical to determining the course of vocal development.[3] This description of speech development is consistent with emerging data and theory concerning vocal learning in songbirds (Konishi, 1989)—both include a period of perceptual learning during which auditory communicative signals heard from adult members of the species are stored in memory followed by a period of perceptual-motor practice. On this account, the early articulations of infants are not just random movements of the "speech limbs" with no consequences. Instead, infants are consolidating mapping rules that relate self-produced auditory events to self-produced movement patterns. The resulting information can be used to guide their future behavior; it informs them about what to do with their articulators to achieve a specific auditory target. It will be of considerable interest to investigate the specificity of the auditory–articulatory map, its initial state, and how its configuration changes with age and linguistic experience.

ACKNOWLEDGMENTS

The research reported here, and work contributing to it, was supported by grants from the National Institutes of Health (HD 22514, DC 00520, and HD 18286). The authors are extremely grateful to Erica Stevens and Karen Williams for their help on all phases of this research and also thank Craig Harris for his valuable assistance.

[1]Note that some matrices have expected values less than 5.0 in more than 20% of the cells. Therefore, as recommended by Everitt (1992) and Mehta and Patel (1986), exact probabilities were computed to derive the significance values for the chi-square tests in these cases. StatXact statistical software (Mehta and Patel, 1992) was used.

[2]An interesting question raised by our findings is whether the imitation of vowels in the laboratory is affected by infants' prior experience with ambient speech. Would infants show vocal imitation only for sounds they have had previously experienced? The answer to this will be determined by future studies comparing infant imitation of previously experienced (native-language) speech with their imitation of previously nonexperienced (foreign-language) speech. In demonstrating vocal imitation for native-language sounds in infants under 20 weeks with 15 min of exposure, an experimental technique and baseline data relevant to this question have been established.

[3]In principle these factors are dissociable. One approach involves studies of infants who are tracheostomized and thus aphonic during early infancy. Locke and Pearson (1992) studied such an infant, a girl who was aphonic from 5 to 20 months. They reported that after decannulation the incidence of canonical babbling was near zero despite normal hearing.

Bosma, J. F. (**1975**). "Anatomic and physiologic development of the speech apparatus," in *The Nervous System*, Vol. 3, Human Communication and its Disorders, edited by D. B. Tower (Raven, New York), pp. 469–480.

Boysson-Bardies, B. de, Halle, P., Sagart, L., and Durand, C. (**1989**). "A crosslinguistic investigation of vowel formants in babbling," J. Child Lang. **16**, 1–17.

Boysson-Bardies, B. de, Halle, P., Sagart, L., and Durand, C. (**1984**). "Discernible differences in the babbling of infants according to target language," J. Child Lang. **11**, 1–15.

Boysson-Bardies, B. de, Vihman, M. M., Roug-Hellichius, L., Durand, C., Landberg, I., and Arao, F. (**1992**). "Material evidence of infant selection from the target language: A cross-linguistic phonetic study," in *Phonological Development: Models, Research, Implications*, edited by C. A. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 369–391.

Everitt, B. S. (**1992**). *The Analysis of Contingency Tables* (Chapman and Hall, New York), 2nd ed., Vol. 45.

Grieser, D., and Kuhl, P. K. (**1989**). "Categorization of speech by infants: Support for speech-sound prototypes," Devel. Psych. **25**, 577–588.

Holmgren, K., Lindblom, B., Aurelius, G., Jalling, B., and Zetterström, R. (**1986**). "On the phonetics of infant vocalization," edited by B. Lindblom and R. Zetterström *Precursors of Early Speech* (Stockton, New York).

Iverson, P., and Kuhl, P. K. (**1995**). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," J. Acoust. Soc. Am. **97**, 553–562.

Iverson, P., and Kuhl, P. K. (**1996**). "Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/," J. Acoust. Soc. Am. **99**, 1130–1140.

Jakobson, R., Fant, C. G. M., and Halle, M. (**1969**). *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates* (MIT, Cambridge, MA).

Kent, R. D. (**1981**). Articulatory–acoustic perspectives on speech development,'' in *Language Behavior in Infancy and Early Childhood*, edited by R. E. Stark (Elsevier, New York), pp. 105–126.

Kent, R. D. (**1992**). ''The biology of phonological development,'' in *Phonological Development: Models, Research, Implications*, edited by C. A. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 65–90.

Kent, R. D., and Forner, L. L. (**1979**). ''Developmental study of vowel formant frequencies in an imitation task,'' J. Acoust. Soc. Am. **65**, 208–217.

Kent, R. D., and Murray, A. D. (**1982**). ''Acoustic features of infant vocalic utterances at 3, 6, and 9 months,'' J. Acoust. Soc. Am. **72**, 353–365.

Kent, R. D., Osberger, M. J., Netsell, R., and Hustedde, C. G. (**1987**). ''Phonetic development in identical twins differing in auditory function,'' J. Speech Hearing Disorders **52**, 64–75.

Kessen, W., Levine, J., and Wendrich, K. A. (**1979**). ''The imitation of pitch in infants,'' Infant Behav. Devel. **2**, 93–99.

Konishi, M. (**1989**). ''Birdsong for neurobiologists,'' Neuron **3**, 541–549.

Koopmans-van Beinum, F. J., and van der Stelt, J. M. (**1986**). ''Early stages in the development of speech movements,'' in *Precursors of Early Speech*, edited by B. Lindblom and R. Zetterström (Stockton, New York).

Kuhl, P. K. (**1979**). ''Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories,'' J. Acoust. Soc. Am. **66**, 1668–1679.

Kuhl, P. K. (**1983**). ''Perception of auditory equivalence classes for speech in early infancy,'' Infant Behav. Devel. **6**, 263–285.

Kuhl, P. K. (**1985**). ''Categorization of speech by infants,'' in *Neonate Cognition: Beyond the Blooming, Buzzing Confusion*, edited by J. Mehler and R. Fox (Erlbaum, Hillsdale, NJ), pp. 231–262.

Kuhl, P. K. (**1991**). ''Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not,'' Percept. Psychophys. **50**, 93–107.

Kuhl, P. K. (**1992**). ''Infants' perception and representation of speech: Development of a new theory,'' in *Proceedings of the International Conference on Spoken Language Processing*, edited by J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, and G. E. Wiebe (University of Alberta Press, Edmonton), pp. 449–456.

Kuhl, P. K. (**1993a**). ''Developmental speech perception: Implications for models of language impairment,'' in *Temporal Information Processing in the Nervous System*, edited by P. Tallal, A. M. Galaburda, R. R. Llinás, and C. von Euler (The New York Academy of Sciences, New York), Vol. 682, pp. 248–263.

Kuhl, P. K. (**1993b**). ''Innate predispositions and the effects of experience in speech perception: The Native-Language Magnet theory,'' in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (Kluwer Academic, Dordrecht, The Netherlands), pp. 259–274.

Kuhl, P. K. (**1994**). ''Learning and representation in speech and language.'' Curr. Opinion Neurobiol. **4**, 812–822.

Kuhl, P., and Iverson, P. (**1995**). ''Linguistic experience and the 'perceptual magnet effect,' '' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 121–154.

Kuhl, P. K., and Meltzoff, A. N. (**1982**). ''The bimodal perception of speech in infancy,'' Science **218**, 1138–1141.

Kuhl, P. K., and Meltzoff, A. N. (**1984**). ''The intermodal representation of speech in infants,'' Infant Behav. Devel. **7**, 361–381.

Kuhl, P. K., Meltzoff, A. N. (**1988**). ''Speech as an intermodal object of perception,'' in *Perceptual Development in Infancy: The Minnesota Symposia on Child Psychology*, edited by A. Yonas (Erlbaum, Hillsdale, NJ), Vol. 20, pp. 235–266.

Kuhl, P. K., and Meltzoff, A. N. (**in press**). ''Evolution, nativism, and learning in the development of language and speech,'' in *The Inheritance and Innateness of Grammars*, edited by M. Gopnik (Oxford U.P., New York), pp. 7–44.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). ''Linguistic experience alters phonetic perception in infants by 6 months of age,'' Science **255**, 606–608.

Legerstee, M. (**1990**). ''Infants use multimodal information to imitate speech sounds,'' Infant Behav. Devel. **13**, 343–354.

Lieberman, P. (**1984**). *The Biology and Evolution of Language* (Harvard U.P., Cambridge, MA).

Lieberman, P. (**1991**). *Uniquely Human: The Evolution of Speech, Thought, and Selfless Behavior* (Harvard U.P., Cambridge, MA).

Lieberman, P., Crelin, E. S., and Klatt, D. H. (**1972**). ''Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee,'' Am. Anthropol. **74**, 287–307.

Locke, J. L. (**1993**). *The Child's Path to Spoken Language* (Harvard U.P., Cambridge, MA).

Locke, J. L., and Pearson, D. M. (**1992**). ''Vocal learning and the emergence of phonological capacity: A neurobiological approach,'' in *Phonological Development: Models, Research, Implications*, edited by C. A. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 91–129.

Marean, G. C., Werner, L. A., and Kuhl, P. K. (**1992**). ''Vowel categorization by very young infants,'' Devel. Psych. **28**, 396–405.

Marler, P. (**1974**). ''Constraints on learning: Development of bird song,'' in *Ethology and Psychiatry*, edited by N. F. White (University of Toronto Press, Toronto), pp. 69–83.

Mehta, C., and Patel, N. (**1992**). *StatXact: Statistical Software for Exact Nonparametric Inference* (CYTEL Corporation, Cambridge, MA).

Mehta, C. R., and Patel, N. R. (**1986**). ''A hybrid algorithm for Fisher's exact test in unordered r×c contingency tables,'' Commun. Statist. **15**, 387–403.

Meltzoff, A. N. (**1990**). ''Towards a developmental cognitive science: The implications of cross-modal matching and imitation for the development of representation and memory in infancy,'' in *Annals of the New York Academy of Sciences*, Vol. 608: The Development and Neural Bases of Higher Cognitive Functions, edited by A. Diamond (New York Academy of Sciences, New York), pp. 1–31.

Meltzoff, A. N., and Kuhl, P. K. (**1994**). ''Faces and speech: Intermodal processing of biologically-relevant signals in infants and adults,'' in *The Development of Intersensory Perception: Comparative Perspectives*, edited by D. J. Lewkowicz and R. Lickliter (Erlbaum, Hillsdale, NJ), pp. 335–369.

Meltzoff, A. N., and Moore, M. K. (**1977**). ''Imitation of facial and manual gestures by human neonates,'' Science **198**, 75–78.

Meltzoff, A. N., and Moore, M. K. (**1983a**). ''Newborn infants imitate adult facial gestures,'' Child Devel. **54**, 702–709.

Meltzoff, A. N., and Moore, M. K. (**1983b**). ''The origins of imitation in infancy: Paradigm, phenomena, and theories,'' in *Advances in Infancy Research*, edited by L. P. Lipsitt (Ablex, Norwood, NJ), Vol. 2, pp. 265–301.

Meltzoff, A. N., and Moore, M. K. (**1992**). ''Early imitation within a functional framework: The importance of person identity, movement, and development,'' Infant Behav. Devel. **15**, 479–505.

Meltzoff, A. N., and Moore, M. K. (**1994**). ''Imitation, memory, and the representation of persons,'' Infant Behav. Devel. **17**, 83–99.

Miller, J. D. (**1989**). ''Auditory-perceptual interpretation of the vowel,'' J. Acoust. Soc. Am. **85**, 2114–2134.

Nearey, T. M. (**1989**). ''Static, dynamic, and relational properties in vowel perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.

Nottebohm, F. (**1975**). ''A zoologist's view of some language phenomena with particular emphasis on vocal learning,'' in *Foundations of Language Development*, edited by E. H. Lenneberg and E. Lenneberg (Academic, New York), Vol. 1, pp. 61–103.

Oller, D. K. (**1978**). ''Infant vocalization and the development of speech,'' Allied Health Behav. Sci. **1**, 523–549.

Oller, D. K., and Eilers, R. E. (**1988**). ''The role of audition in infant babbling,'' Child Devel. **59**, 441–449.

Oller, D. K., Eilers, R. E., Bull, D. H., and Carney, A. E. (**1985**). ''Pre-speech vocalizations of a deaf infant: A comparison with normal metaphonological development,'' J. Speech Hearing Res. **28**, 47–63.

Oller, D. K., and Lynch, M. P. (**1992**). ''Infant vocalizations and innovations in infraphonology: Toward a broader theory of development and disorders,'' in *Phonological Development: Models, Research, Implications*, edited by C. A. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 509–536.

Papousek, M., and Papousek, H. (**1981**). ''Musical elements in the infant's vocalization: Their significance for communication, cognition, and creativity,'' in *Advances in Infancy Research*, edited by L. P. Lipsitt and C. K. Rovee-Collier (Ablex, Norwood, NJ), Vol. 1, pp. 164–224.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., and Jordan, M. I. (**1993**). ''Trading relations between tongue-body raising and lip rounding in pro-

duction of the vowel /u/: A pilot 'motor equivalence' study,'' J. Acoust. Soc. Am. **93**, 2948–2961.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Piaget, J. (**1962**). *Play, Dreams and Imitation in Childhood* (Norton, New York).
Roug, L., Landberg, I., and Lundberg, L. J. (**1989**). ''Phonetic development in early infancy: A study of four Swedish children during the first eighteen months of life,'' J. Child Lang. **16**, 19–40.
Sasaki, C. T., Levine, P. A., Laitman, J. T., and Crelin, E. S. (**1977**). ''Postnatal descent of the epiglottis in man,'' Arch. Otolaryngol. **103**, 169–171.
Stark, R. E. (**1980**). ''Stages of speech development in the first year of life,'' in *Child Phonology: Vol. 1, Production*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York), Vol. 1, pp. 73–92.
Stark, R. E. (**1983**). ''Phonatory development in young normally hearing and hearing impaired children,'' in *Speech of the Hearing Impaired: Research, Training, and Personnel Preparation*, edited by I. Hochberg, H. Levitt, and M. J. Osberger (University Park, Baltimore), pp. 251–266.
Stoel-Gammon, C. (**1988**). ''Prelinguistic vocalizations of hearing-impaired and normally hearing subjects: A comparison of consonantal inventories,'' J. Speech Hearing Disorders **53**, 302–315.
Stoel-Gammon, C. (**1992**). ''Prelinguistic vocal development: Measurement and predictions,'' in *Phonological Development: Models, Research, Implications*, edited by C. A Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 439–456.
Stoel-Gammon, C., and Cooper, J. A. (**1984**). ''Patterns of early lexical and phonological development,'' J. Child Lang. **11**, 247–271.
Stoel-Gammon, C., and Otomo, K. (**1986**). ''Babbling development of hearing-impaired and normally hearing subjects,'' J. Speech Hearing Disorders **51**, 33–41.
Stoel-Gammon, C., Williams, K., and Buder, E. (**1994**). ''Cross-language difference in phonological acquisition: Swedish and American /t/,'' Phonetica **51**, 146–158.
Studdert-Kennedy, M. (**1986**). ''Development of the speech perceptuomotor system,'' in *Precursors of Early Speech*, edited by B. Lindblom and R. Zetterström (Stockton, New York), pp. 205–217.
Studdert-Kennedy, M. (**1993**). ''Some theoretical implications of cross-modal research in speech perception,'' in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (Kluwer Academic, Dordrecht, The Netherlands), pp. 461–466.
Syrdal, A. K., and Gopal, H. S. (**1986**). ''A perceptual model of vowel recognition based on the auditory representation of American English vowels,'' J. Acoust. Soc. Am. **79**, 1086–1100.
Vihman, M. M., and Miller, R. (**1988**). ''Words and babble at the threshold of language acquisition,'' in *The Emergent Lexicon: The Child's Development of a Linguistic Vocabulary*, edited by M. D. Smith and J. L. Locke (Academic, New York), pp. 151–183.
Walton, G. E., Shoup-Pecenka, A. G., and Bower, T. G. R. (**1991**). ''Speech categorization in infants: Newborns can detect phonetic invariance across talkers,'' J. Acoust. Soc. Am. **90**, 2296 (A).