VOCAL LEARNING IN INFANTS: DEVELOPMENT OF PERCEPTUAL-MOTOR LINKS FOR SPEECH

Patricia K. Kuhl and Andrew N. Meltzoff
University of Washington, Seattle, WA

ABSTRACT

Infants' development of speech begins with a language-universal pattern of production that eventually becomes language-specific. Our work focuses on the processes underlying this developmental change in infants' production of speech. One important contribution to this change is vocal learning — the process by which infants listen to the patterns of ambient language and attempt to reproduce them. We here discuss studies on 12- to 20-week-old infants' vocalizations in response to speech. The results show that (a) very young infants imitate sound patterns they hear, and (b) developmental change occurs in the microstructure of infants' vowel categories. We conclude that connections between auditory and articulatory representations of speech exist very early in life.

INTRODUCTION

Speech production during the first two years of life has been described by universal stages [1]. These changes encompass: *Reflexive Phonation* (0 to 2 months), *Cooing* (1 to 4 months), *Expansion* (3 to 8 months), *Canonical Babbling* (5 to 10 months), and *Meaningful Speech* (10 to 18 months). Although there is considerable consensus on the description of successive stages, there is less known about the *processes* by which change in infants' vocalizations are induced. Two factors are critical in producing the transition in infants, anatomical change and vocal learning. The young infant's vocal tract is very different from that of the adult, and anatomical growth must contribute, at least in part, to the stage-like shift seen in infants' vocalizations.

The factor emphasized here, however, is *vocal learning*. Human infants learn speech by listening to ambient language and attempting to produce sound patterns that match what they hear [2]. Cross-cultural studies show that infants from different linguistic environments exhibit differences in their vocalizations by 1 year of age, suggesting that by this age, ambient language has had an effect on spontaneous speech production [3].

The question examined here is whether there is evidence of vocal learning earlier in life. One way to investigate this point is to study vocal imitation. It offers a behavioral assay of vocal learning. Vocal imitation requires that infants recognize the relationship between articulatory movements and sound. 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 Evidence

Infants' vocalizations in response to speech were recorded at three ages, 12-, 16-, and 20-weeks of age [4]. Infants (N = 72) watched and listened to a female producing one of three vowels, /a/, /i/, or /u/. The vowels were produced once every 5 sec by a female talker and presented via a video display (life-size and in color). The stimuli were presented to infants as they reclined in an infant seat at a distance of 18" from the monitor.

All of the infants' vocalizations were recorded. Those meeting pre-established criteria for "vowel-like" utterances (greater than 100 msec and fully resonant, N = 224) were analyzed perceptually by having them phonetically transcribed, and analyzed instrumentally using computerized spectrographic techniques. The phonetic transcriber used a narrow transcription to initially code the vowels: /i, I, ɛ, ae, a, ʌ, U, u/. Once transcribed, the vowels were classified into one of three groups: /a/-like vowels (/ae/, /a/, /ʌ/), /i/-like vowels /i/, /I/, /ɛ/), and /u/-like vowels (/U/, /u/). A second individual rescored 30% of the utterances, chosen randomly. Transcriptional reliability for the three-category coding was 92%.

Instrumental analysis was conducted by a person who was blind to the transcriptional classification of each utterance. Infants' vocalizations were analyzed using Kay Elemetrics

microcomputer-based equipment (CSL, v. 4.0). The first two formants of each utterance were sampled at five locations: onset, at the 1/4, 1/2, and 3/4 points, and offset.

Two outcomes of the study are noteworthy for theory. First, there was developmental change in infants' vowel productions. Figure 1 displays the /a/-like, /i/-like, and /u/-like vowels of 12-, 16-, and 20-week-old infants in an F1/F2 coordinate space. In each graph, infants' vowel utterances are classified according to the phonetic transcription data. The closed circles enclose 90% of the utterances in each category. As shown, utterances in each of the three categories formed clusters in acoustic space. These clusters are, from an acoustic standpoint, relationally consistent with productions by adult speakers.

More intriguing from a developmental standpoint, the areas of vowel space occupied by infants' /a/, /i/, and /u/ vowels become progressively more *separated* between 12- and 20-weeks of age. Infant vowel categories were more tightly clustered at 20 weeks than at 12 weeks. What causes the increased separation of vowel categories over this relatively short (8 week) period? We suggest that infants listening to their ambient language have begun to form *stored representations* of vowels in memory; these representations serve as "targets" that infants try to match. Our view is that the stored representations resulting from infants' analysis of ambient language influences not only their perception of speech [see 11 for discussion] but their subsequent productions as well.

A second result of the study suggested that infants' vowel productions can be influenced by short-term exposure to sound. Infants' vowel productions were altered depending on what they heard. Infants produced more /a/-like utterances when exposed to /a/ than when exposed to /i/ or /u/; similarly, they produced more /i/-like utterances when exposed to /i/ than when exposed to /a/ or /u/; finally, they produced more /u/-like utterances when exposed to /u/ than when exposed to /a/ or /i/. In sum, the particular vowel stimulus infants heard influenced the type of vowel infants produced.

The total amount of exposure that infants received to their target vowel was only 15 minutes (5-min exposure to a specific vowel for each of three days). If 15 minutes of laboratory exposure influences infants' vocalizations, then listening to the ambient language for 20 weeks would appear to provide sufficient exposure to induce change.

How do 12 week-old-infants know how to move their articulators in a way that achieves a specific auditory target? Research on visually-specified movements shows that newborns imitate silent movements of the articulators, such as mouth opening and tongue protrusion [5]. We do not know if this kind of mapping from auditory to articulatory events exists at birth, but it is not out of the question, given Meltzoff and Moore's findings of visual-motor mappings of human mouth movements.

Even if primitive connections exist initially, they must be rapidly expanded to create the repertoire that infants possess just a short time later. This rapid expansion is gained, we believe, through

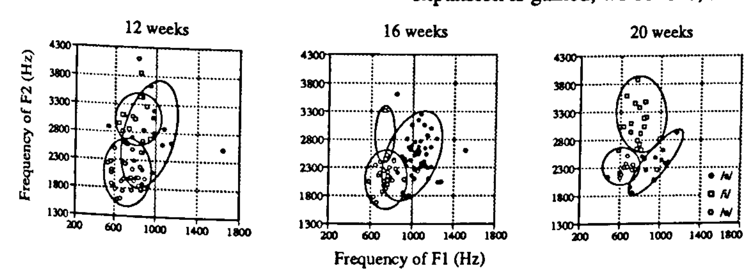

Figure 1. The location of /a/, /i/, and /u/ vowels produced by 12-, 16-, and 20-week-old infants. The curves were drawn by visual inspection to enclose 90% or more of the infants' utterances. Across age, infants' vowel productions show tighter clustering in vowel space. From Kuhl & Meltzoff (In press).

experience as infants engage in cooing and sound play. Infant cooing, which begins at about 4 weeks of age, allows extensive exploration of the nascent auditory-articulatory "map" during which (self-produced) auditory events are related to the motor movements that caused them. Presumably, infants' accuracy in producing vowels improves as infants relate the acoustic consequences of their own articulatory acts to the acoustic targets they heard. This implies that infants not only have to hear the sounds produced by others, but hear the results of their own attempts to speak. Hearing the sound patterns of ambient language (auditory exteroception) as well as hearing one's own attempts at speech (auditory proprioception) are critical to determining the course of vocal development.

Polymodal Speech Representations

Research on adults shows that speech perception is a polymodal phenomenon in which vision plays a role. This is indicated by auditory-visual "illusions" that result when discrepant information is sent to the two separate modalities [6,7,8].

Even very young infants appear to represent speech polymodally. We demonstrated that 18-20-week-old infants recognize auditory-visual connections, akin to what we as adults do when we lipread [9]. Four-month-old infants viewed two filmed faces, side by side, of a woman pronouncing two vowels silently and in synchrony, the vowel /a/ and the vowel /i/. Infants heard one of the two vowels (either /a/ or /i/), played in synchrony with the faces from a loudspeaker located midway between the two faces. The results of the test showed that infants who heard the vowel /a/ looked longer at the face pronouncing /a/ while the infants who heard /i/ looked longer at the vowel /i/. The experiment shows that by four months infants can recognize that an /a/ sound is mapped to a face articulating /a/ while an /i/ sound is mapped to an /i/ articulation. Infants' cross-modal speech perception abilities provide direct evidence that infants connect sound and articulatory movement when they observe another person speak.

Developmental Speech Theory

We would like to offer a more general developmental model about the connection between audition and articulation early in infancy. We emphasize two terms: representation and the influence of ambient language. Ambient language comes into play because it provides input to the child that shapes his or her perceptual space even before words are understood or produced [11,12]. Representation comes into play because we also believe that these speech patterns are stored in memory — that they are represented by the child and serve as "targets" which infants try to match when they vocalize.

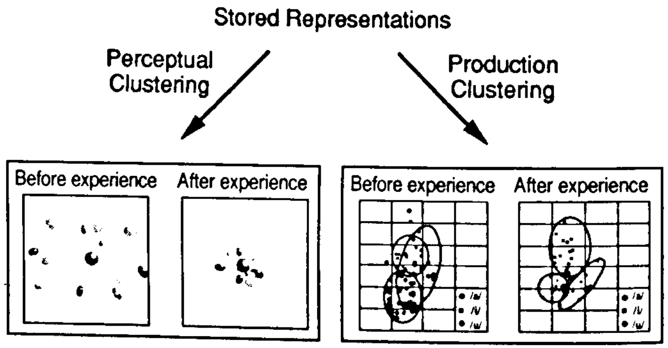

Figure 2. Infants' stored representations affect both speech perception, producing the perceptual clustering evidenced by the magnet effect, as well as speech production, producing the increased clustering seen in infants' vocalizations.

This framework is useful in interpreting the production study reported here. The study uncovered both long-term and short-term changes in infants' vocal repertoires. Long-term changes occurred over the eight-week period (from 12- to 20-weeks) during which infants' vocalizations were measured); short-term changes occurred in infants' vocalizations with relatively short periods of laboratory exposure that were manifest in vocal imitation. These findings suggest that infants' acquisition of speech is strongly influenced by the auditory information that surrounds them.

Two lines of research converge in speech development. On the one hand, there are the findings that linguistic exposure alters infants' *perception* of speech [11,12], and on the other hand, the result shown here that linguistic exposure alters infants' *production* of speech. These can be unified by the suggestion that stored representations of speech alter current processing, and that this occurs from the very earliest phases of infancy. On this view, the tighter clustering observed in infant vowel production and the tighter clustering among vowels in infant perception are both attributable to a common underlying mechanism — the *formation of representations that derive initially from perception of the ambient input and then act as targets for motor output.* The speech representational system is thus deeply and thoroughly polymodal. Early listening affects both sensory perception and motor learning.

ACKNOWLEDGMENT

Research reported here was supported by grants from NIH (HD-22514, HD-18286, and DC 00520).

REFERENCES

[1] Stoel-Gammon, C. (1992), "Prelinguistic vocal development: Measurement and predictions", in *Phonological development: Models, research, implications*, edited by Ferguson, C.A., Menn, L., Stoel-Gammon, C., pp. 439-456. Timonium, MD: York.

[2]. Kuhl, P. K., Meltzoff, A. N. (In press), "Evolution, nativism, and learning in the development of language and speech", in *The biological basis of language*, edited by Gopnik, M., New York: Oxford University Press.

[3] de Boysson-Bardies, B. (1993), "Ontogeny of language-specific syllabic productions", in *Developmental neurocognition: Speech and face processing in the first year of life*, edited by de Boysson-Bardies, B., de Schonen, S., Jusczyk, P., McNeilage, P., Morton, J., pp. 353-363. Dordrecht, Netherlands: Kluwer.

[4] Kuhl, P. K., Meltzoff, A. N. (In press), "Infant vocalizations in response to speech: Vocal imitation and developmental change", *Journal of the Acoustical Society of America.*

[5] Meltzoff, A. N., Moore, M. K. (1994), "Imitation, memory, and the representation of persons", *Infant Behavior and Development*, vol. 17, pp. 83-99.

[6] Green, K. P., Kuhl, P. K., Meltzoff, A. N., Stevens, E. B. (1991), "Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect", *Perception & Psychophysics*, vol. 50, pp. 524-536.

[7] Massaro, D. W., Cohen, M. M. (1994), "Auditory/visual speech in multimodal human interfaces", in *Proceedings of the International Conference on Spoken Language Processing*, pp. 531-534. Tokyo: Acoustical Society of Japan.

[8] Kuhl, P. K., Tsuzaki, M., Tohkura, Y., Meltzoff, A. N. (1994), "Human processing of auditory-visual information in speech perception: Potential for multimodal human-machine interfaces", in *Proceedings of the International Conference on Spoken Language Processing*, pp. 539-542. Tokyo: Acoustical Society of Japan.

[9] Kuhl, P. K., Meltzoff, A. N. (1982), "The bimodal perception of speech in infancy", *Science*, vol. 218, pp. 1138-1141.

[10] Kuhl, P. K. (1994), "Learning and representation in speech and language", *Current Opinion in Neurobiology*, vol. 4, pp. 812-822.

[11] Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., Lindblom, B. (1992), "Linguistic experience alters phonetic perception in infants by 6 months of age", *Science*, vol. 255, pp. 606-608.

[12] Kuhl, P. K. (this volume), "Mechanisms of developmental change in speech and language".