Research Report

Infant Rule Learning Facilitated by Speech

Gary F. Marcus, Keith J. Fernandes, and Scott P. Johnson

New York University

ABSTRACT—Sequences of speech sounds play a central role in human cognitive life, and the principles that govern such sequences are crucial in determining the syntax and semantics of natural languages. Infants are capable of extracting both simple transitional probabilities and simple algebraic rules from sequences of speech, as demonstrated by studies using ABB grammars (la ta ta, gai mu mu, etc.). Here, we report a striking finding: Infants are better able to extract rules from sequences of nonspeech— such as sequences of musical tones, animal sounds, or varying timbres—if they first hear those rules instantiated in sequences of speech.

A hallmark of human language is its abstract character; learners do not simply memorize particular sentences, but rather learn generalizable rules that govern the sequencing of linguistic elements, both familiar and unfamiliar. Proceeding from recent observations that infants are able to extract transitional probabilities from both speech sequences (Saffran, Aslin, & Newport, 1996) and nonspeech sequences (e.g., musical tones: Saffran, Johnson, Aslin, & Newport, 1999; visual shapes: Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002), we exposed infants to simple algebraically structured temporal sequences (Marcus, Vijayan, Bandi Rao, & Vishton, 1999) consisting of either speech syllables (naturally sung or synthesized) or nonspeech sounds (pure tones, instrument timbres, and animal sounds) and asked whether infants could extract rules from those sequences. To our surprise, in our first two experiments, infants were (at least under conditions matched to the original speech studies of Marcus et al., 1999) able to do so only when exposed to speech sequences. In a third experiment, we discovered an intriguing result: If infants first heard a rule instantiated over speech, they were able to generalize that rule to sequences of nonspeech

Address correspondence to Gary F. Marcus, Department of Psychology, 6 Washington Place, New York University, New York, NY 10003, e-mail: gary.marcus@nyu.edu.

sounds, an instance in which transfer (to a new domain) appears to be easier than learning directly (from that novel domain).

GENERAL METHOD

Subjects

The final sample comprised one hundred twenty-eight 7.5-month-old infants (mean age = 226.6 days, SD = 20.8; 53 girls, 75 boys). Sixteen infants were observed in each condition (32 in Experiment 1, 48 in Experiment 2, 48 in Experiment 3). An additional 45 infants were excluded because of fussiness (19), inattention (12), technical difficulties (4), experimenter error (8), or parental interference (2). All infants were full-term with no known developmental difficulties. They were recruited from a commercial database of new parents.

Procedure

Infants sat in a parent's lap, in a testing chamber with speakers and visual accompaniment on either side. They were familiarized with three presentations each of 16 different sequences following a particular grammar (ABA, ABB, or AAB), with a 0.25-s gap between individual sounds and a 1-s gap between sequences. Sequences were presented from both sides, at equal volume. Familiarization time was 2.38 min in Experiment 1 and 2 min in Experiments 2 and 3. After familiarization, infants were tested on four unique sequences composed of new items. Two of these new sequences were consistent with the familiarization grammar, and two were inconsistent with it. Each pair was repeated for up to 15 s per trial for a total of 12 test trials. Each test sequence was presented on either the left or the right side and paired with a small flashing light (or a computer-generated, brightly colored spinning wheel in the case of timbre sounds in Experiment 2) on the same side. Inconsistent and consistent sequences were presented in alternation. Sung syllables in Experiment 1 were approximately 500 ms in duration; all other sounds were approximately 300 ms long.

The dependent measure was the amount of time spent looking in the direction of the light (or monitor) associated with the

speaker presenting a given test sequence. Familiarization rule (ABB vs. ABA or ABB vs. AAB), order of presentation of the test sequences (consistent vs. inconsistent first), and side of presentation (left vs. right side first at test) were counterbalanced across infants in each condition. There were no reliable effects of these variables, or of test trial, in preliminary analyses.

Apparatus and Stimuli

A Macintosh G4 computer and Radio Shack speakers (AMX-7) were used to present stimuli and collect looking-time data. An observer viewed the infant on a monitor and held down a key whenever the infant was fixated on the visual stimulus. The observer and parent wore headphones and listened to masking music, and were unaware of the stimulus patterns presented to the infant at test. On each trial, the stream of test sequences was presented as long as the infant attended to the visual stimulus. A trial ended when the infant looked away for 2 s, or had looked for 15 s.

EXPERIMENT 1

In the first experiment, we examined infants' rule learning using sequences of tones and sung syllables that matched the tones in pitch.

Stimuli

In the *tones condition*, the 16 familiarization sequences were produced with piano notes produced by Easy Beat software in the octave beginning with middle C on a standard keyboard (A = 440 Hz). The ABB sequences were CFF, CD#D#, CEE, CDD, C#FF, C#D#D#, C#EE, C#DD, F#FF, F#D#D#, F#EE, F#DD, GFF, GD#D#, GEE, and GDD. AAB items were constructed with the same inventory of A items (C, C#, F#, and G) and B items (F, D#, E, and D). Test sequences were G#AA and BA#A# (consistent with ABB) and G#G#A and BBA# (consistent with AAB). Because the contours varied across items, an infant could not succeed by transposing a single melody; instead, success required extraction of the AAB or ABB pattern away from the varying individual contours. The intervals in familiarization sequences ranged from 1 to 5 semitones (mean of 3), and the intervals in the test sequences were always 1 semitone (i.e., the distance between C and C#).[1]

In the *sung-syllables condition*, the 16 familiarization sentences following the ABB pattern (as in Marcus et al., 1999) were *le di di, le je je, le li li, le we we, wi di di, wi je je, wi li li, wi we we, ji di di, ji je je, ji li li, ji we we, de di di, de je je, de li li,* and *de we we* (per IPA notation, /i/ and /e/ represent the vowel sounds of the

words *meet* and *mate*, respectively, in standard American English); AAB items were constructed with the same vocabulary of A items (*le, wi, ji,* and *de*) and B items (*di, je, li,* and *we*). The test sequences were *ba po po* and *ko ga ga* (consistent with ABB) and *ba ba po* and *ko ko ga* (consistent with AAB). Each syllable was sung at a particular note matching a pitch in the tones condition: *le* at C, *wi* at C#, *ji* at F#, *de* at G, *di* at F, *je* at D#, *li* at E, *we* at D, *ba* at G#, *po* at A, *ko* at B, and *ga* at A#. The stimuli were produced by a musically trained female vocalist and were recorded and edited using Sound Studio software.

Results and Discussion

Looking times were entered into a 2 (condition: tones vs. sung syllables) × 2 (test pattern: consistent vs. inconsistent) mixed analysis of variance (ANOVA), which yielded a reliable main effect of test pattern, $F(1, 30) = 5.67, p = .024, p_{rep} = .919, d = 0.416$, and a reliable Condition × Test Pattern interaction, $F(1, 30) = 4.99, p = .033, p_{rep} = .903$. Simple effects tests revealed a preference for the inconsistent pattern in the sung-syllables condition, $F(1, 30) = 10.65, p = .0028, p_{rep} = .975, d = 0.886$, but no significant preference in the tones condition, $F(1, 30) = 0.01$, n.s. (see Fig. 1).

In our tones task—which demanded that infants go beyond specific melodic contours toward recognition of abstract reduplicative structures—infants showed no reliable difference in looking time for consistent versus inconsistent test patterns. Thus, although infants can recognize transpositions in contour, which are perhaps recoded as instantiations of a common melody (Trehub, Bull, & Thorpe, 1984; Trehub, Thorpe, & Morrongiello, 1987), the infants in Experiment 1 were unable to detect the abstract underlying reduplicative pattern in familiarization strings (consisting of a mixture of strings, some with rising contours and others with falling contours). In contrast, given sequences of syllables sung at pitches identical to those in the tones condition, infants subsequently listened longer to sequences with a different pattern; this finding suggests that infants' mechanisms for extracting rules from sequential auditory materials, unlike their mechanisms for transitional-probability learning, may be facilitated by phonetic content.

EXPERIMENT 2

To address the possibility that the materials in Experiment 1 might have been insufficiently engaging or difficult to encode, we investigated whether infants could extract rules from sequences of musical sounds differing only in timbre (the *Timbre 1 condition*), an acoustic property that infants are known to encode and remember reliably (Trainor, Wu, & Tsang, 2004). We then replicated this condition using a slightly different grammatical contrast (the *Timbre 2 condition*), testing ABB versus ABA. Finally, both as a test of the role of salience and to guard against the possibility that synthesized sounds were somehow

---

[1]Although these intervals are well within the range of discriminability for infants (Trehub, Bull, & Thorpe, 1984), we conducted a replication experiment with wider intervals, both in familiarization (2–14 semitones, mean of 8) and in test (6 and 8 semitones). Results were consistent with those reported in the main text: We found no reliable difference in preference for consistent versus inconsistent test sequences, $t(15) = 0.22$, n.s.
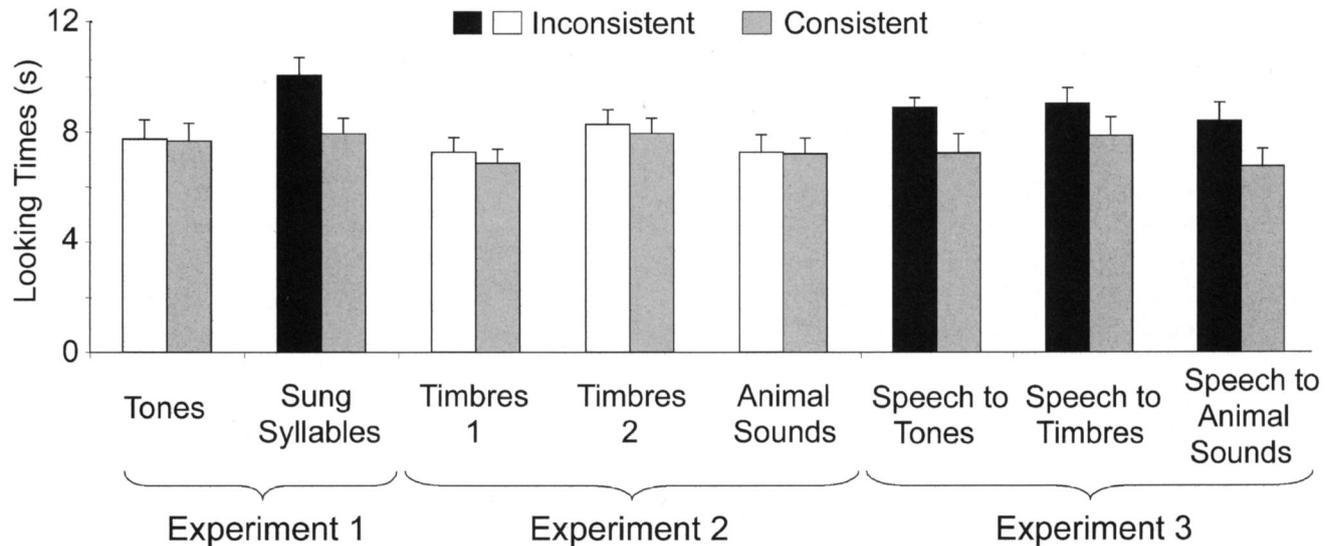
Fig. 1. Mean looking times for inconsistent and consistent test items in each of the eight conditions. Light-gray bars denote looking times to consistent items; black bars denote looking times to inconsistent items following exposure to speech, and white bars denote looking times to inconsistent items following exposure to nonspeech sequences. Error bars indicate standard errors of the means.

ecologically inappropriate, we ran an additional condition using sequences of recorded animal sounds (the *animal-sounds condition*).

Stimuli

The timbre stimuli, generated by a Roland JV-90 Multitimbral Synthesizer, were piano, violin, French horn, clarinet, trumpet, flute, vibraphone, harp, saxophone, oboe, organ, and chime voiced at 262 Hz (middle C). The animal sounds, gathered from the Internet (available on request), were recordings of live animals: cat, crow, goat, hen, dog, duck, sheep, sparrow, fox, puppy, sea lion, and whale. Familiarization and test sequences followed the same grammars as in Experiment 1, with the exception that the Timbre 2 condition contrasted ABB with ABA.

Results and Discussion

Looking times were entered into a 3 (condition: Timbre 1 vs. Timbre 2 vs. animal sounds) × 2 (test pattern: consistent vs. inconsistent) mixed ANOVA, which yielded no reliable effects (see Fig. 1). Again, therefore, infants were unable to detect patterns in streams of nonlinguistic stimuli. This experiment further supports the conclusion that sequences of nonspeech stimuli do not readily engage infants' machinery for extracting rules.

EXPERIMENT 3

Against this background, and drawing on a theoretically motivated distinction between rule extraction (or discovery) and generalization (or extension; Anderson, 1983), we asked whether an infant who has acquired a rule from speech can subsequently generalize that rule to a new domain. Experiment 3 included three crossover conditions, with tones, timbres, and animal sounds. Infants were familiarized with structured sequences of speech (ABB or ABA, between subjects) and then tested on their ability to discriminate those same structures instantiated in sequences of nonspeech stimuli. Looking times to consistent and inconsistent test stimuli were compared in a *speech-to-tones condition*, a *speech-to-timbres condition*, and a *speech-to-animal-sounds condition*.

Stimuli

The familiarization stimuli (computer-generated spoken syllables) were identical to those employed by Marcus et al. (1999); the test stimuli were identical to those employed in Experiment 2 of the present report.

Results and Discussion

Looking times were entered into a 3 (condition: speech to tones vs. speech to timbres vs. speech to animal sounds) × 2 (test pattern: consistent vs. inconsistent) mixed ANOVA, which yielded a reliable main effect of test pattern, $F(1, 45) = 19.96$, $p < .001$, $p_{rep} = .997$, $d = 0.579$, and no other significant effects (see Fig. 1). Separate planned comparisons (simple-effects tests) revealed significant preferences for the inconsistent pattern in all three conditions: speech to tones, $F(1, 45) = 9.21$, $p = .004$, $p_{rep} = .970$, $d = 0.745$; speech to timbres, $F(1, 45) = 6.51$, $p = .014$, $p_{rep} = .940$, $d = 0.525$; speech to animal sounds, $F(1, 45) = 4.62$, $p = .037$, $p_{rep} = .897$, $d = 0.459$. All three crossover conditions, therefore, pointed in the same direction:

Prior exposure to ABB or ABA structures in speech facilitated generalization of those structures to sequences of nonlinguistic stimuli.

We conducted additional analyses to compare performance across the three experiments, beginning with a 2 (familiarization stimulus: speech vs. nonspeech) $\times$ 2 (test pattern: consistent vs. inconsistent) mixed ANOVA. This analysis yielded a reliable main effect of familiarization stimulus, $F(1, 126) = 3.92$, $p = .049$, $p_{rep} = .879$, $d = 0.349$, stemming from a novelty preference after listening to speech; a reliable main effect of test pattern, $F(1, 126) = 17.81$, $p < .001$, $p_{rep} = .997$, $d = 0.364$, reflecting a preference for the inconsistent test pattern overall; and a significant interaction, $F(1, 126) = 10.29$, $p = .002$, $p_{rep} = .981$. In the conditions in which infants were familiarized with speech, there was a reliable novelty preference, $F(1, 126) = 27.58$, $p < .001$, $p_{rep} > .999$, $d = 0.640$, whereas in the conditions in which familiarization consisted of nonspeech stimuli, there was no reliable preference, $F(1, 126) = 0.51$, n.s. Additionally, as shown in Figure 1, there was no reliable difference among looking times for consistent items across the speech and nonspeech test stimuli (light-gray bars), $F(1, 126) = 0.01$, n.s.; this result suggests that the diverse materials were of roughly equal intrinsic interest. In contrast, the mean looking time to inconsistent items was markedly greater for infants familiarized with speech (black bars) than for infants familiarized with nonspeech (white bars), $F(1, 126) = 11.18$, $p = .001$, $p_{rep} = .985$, $d = 0.591$. Across the eight conditions as a whole, therefore, the results are clear: Speech is a catalyst for engaging infants' machinery for generalizing rules.

GENERAL DISCUSSION

Our results suggest that speech can facilitate rule learning (or rule generalization) in domains where infants might otherwise not acquire rules. Adult humans can extract statistics and rules from a variety of domains (Altmann, Diennes, & Goode, 1995; though perhaps not equally from all domains—see Conway & Christiansen, 2005), and infants can extract rules from visual stimuli that are presented co-temporally such that the relevant pattern can be apprehended in a single glance (Tyrell, Stauffer, & Snowman, 1991; Tyrell, Zingardo, & Minard, 1993). But extracting abstract structure from sequences that extend over time may be more challenging; presumably, this requires establishing memory traces of individual elements within and across trials (Oakes & Ribar, 2005). Experiments 1 and 2, as well as unpublished data with visual stimuli (Marcus, Fernandes, Johnson, & Slemmer, 2004), suggest that under these more demanding circumstances, in which a pattern cannot be apprehended in a single glance, it is easier to extract rules from speech than from nonspeech stimuli.

We do not yet know why speech is capable of playing the facilitative role we observed in the crossover study (Experiment 3). Infants may analyze speech more deeply than other signals because it is highly familiar or highly salient, because it is produced by humans, because it is inherently capable of bearing meaning, or because it bears some not-yet-identified acoustic property that draws the attention of the rule-induction system. Regardless, from birth, infants prefer listening to speech over listening to closely matched control stimuli (Vouloumanos & Werker, 2004, 2007), and the intriguing patterns we have observed in rule learning and transfer could in some way be an extension of that initial, profound interest in speech.

REFERENCES

Altmann, G.T., Diennes, Z., & Goode, A. (1995). On the modality-independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 899–912.

Anderson, J.R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Conway, C.M., & Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 24–39.

Fiser, J., & Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA, 99,* 15822–15826.

Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83,* B35–B42.

Marcus, G.F., Fernandes, K., Johnson, S.P., & Slemmer, J.A. (2004, November). *What's special about speech?: Evidence from a contrast between rules and statistics.* Paper presented at the Boston University Conference on Language Development, Boston, MA.

Marcus, G.F., Vijayan, S., Bandi Rao, S., & Vishton, P.M. (1999). Rule learning in 7-month-old infants. *Science, 283,* 77–80.

Oakes, L.M., & Ribar, R.J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy, 7,* 85–98.

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

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27–52.

Trainor, L.J., Wu, L., & Tsang, C.D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, *7*, 289–296.

Trehub, S., Bull, D., & Thorpe, L.A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*, 821–830.

Trehub, S., Thorpe, L.A., & Morrongiello, B.A. (1987). Organizational processes in infants' perception of auditory patterns. *Child Development*, *58*, 741–749.

Tyrell, D.J., Stauffer, L.B., & Snowman, L.G. (1991). Perception of abstract identity/difference relationships by infants. *Infant Behavior & Development*, *14*, 125–129.

Tyrell, D.J., Zingardo, M.C., & Minard, K.L. (1993). Learning and transfer of identity-difference relationships by infants. *Infant Behavior & Development*, *16*, 43–52.

Vouloumanos, A., & Werker, J.F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*, 270–276.

Vouloumanos, A., & Werker, J.F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*, 159–164.