PROCEEDINGS OF

THE 25TH ANNUAL

Boston University Conference on Language Development

VOLUME 1

Edited by Anna H.-J. Do,

Laura Domínguez, and Aimee Johansen

Coarticulatory Cues Enhance Infants' Recognition of Syllable Sequences in Speech

Suzanne Curtin, Toben H. Mintz, and Dani Byrd
University of Southern California

1.0 Introduction

Infant speech perception research has provided evidence for infants' sensitivity to multiple aspects of the acoustic signal such as VOT, place contrasts, and various other native and non-native contrasts (Eimas & Miller, 1987; Jusczyk, 1994; Jusczyk, 1997; Jusczyk, Cutler & Redanz, 1993; McQueen, 1998; Morgan & Saffran, 1995; Saffran, Aslin, & Newport, 1996; Werker & Pegg, 1992; and others). Recent work has demonstrated that infants are sensitive to a variety of cues in the speech signal, for instance, transitional probabilities, phonotactics, and stress (Jusczyk, 1997; Jusczyk, Cutler, & Redanz, 1993; McQueen, 1998; Morgan & Saffran, 1995; Saffran, Aslin, & Newport, 1996; Saffran, Newport & Aslin, 1996; Turk, Jusczyk, & Gerken, 1995). These research programs have focused on categorical properties of the language and distributional and statistical properties of the language input. Another type of information, coarticulatory cues, is neither categorical nor statistical in the traditional sense. The present experiment explores whether infants use this kind of gradient information to store and recognize syllable sequences.

Infants appear to be sensitive to coarticulation. Recent work by Johnson and Jusczyk (in press) demonstrates that coarticulatory information is important in early word segmentation. In particular, they found that infants are able to use coarticulatory cues to segment speech, and when pitted against statistical properties of the input—specifically transitional probabilities—the coarticulatory cues override the statistical probabilities.

In addition to segmentation, coarticulation provides contextual information about the sound combinations that occur in syllables. Speech sounds are not produced exactly the same way in every context. Rather, the articulation of a particular sound is affected by the surrounding sounds resulting in coarticulation. For example, in the English words 'key' [ki] and 'coo' [ku], the [k] sounds are produced differently. In the case of 'coo' [ku], there is lip rounding on the [k] in anticipation of the rounded vowel that follows. This is not the case for the word 'key' [ki]. The present experiment explores whether coarticulatory information is an important aspect of infants' memory for syllable sequences. Specifically, we investigated whether coarticulatory cues affect infants' *recognition* of syllable sequences. To address these questions, we conducted an experiment investigating the role of coarticulation information in 7-month old infants' ability to recognize previously familiarized syllable sequences. We examined whether infants are sensitive to the appropriateness of coarticulatory cues by probing whether perceived familiarity of a string suffers when it incorporates inappropriate coarticulatory information. The experiment provides evidence for whether the sequential syllable information is sufficient for the recognition of sequences or whether coarticulatory information also plays an important role. Moreover, the experiment addresses whether conflicting coarticulatory information is detrimental to the recognition of syllable sequences. We suggest that coarticulation is a salient source of contextual information and plays an important role in infants' representations of syllable sequences.

2. Experiment: Coarticulatory cues enhance infants' recognition of syllable sequences.

The experiment uses natural speech stimuli in an artificial-language-learning paradigm. These stimuli are used to test whether having items constructed with appropriately coarticulated syllables sequences rather than miscoarticulated syllable sequences affects infants' ability to recognize familiar syllable sequences. The experiment consists of two parts: a familiarization phase followed by a testing phase. The stimuli materials used will be presented first. Following this description, the experimental procedure will be discussed.

2.1 Method

Materials. The familiarization stimuli consisted of a repeated sequence of 27 syllables with certain controlled properties. First, every third syllable in the familiarization string was stressed, and there were no transitional properties indicative of word boundaries. There are several reasons for incorporating stress into the familiarization phase. Work by Fowler (1981) has demonstrated that the second formant transition (F2) of unstressed vowels is affected by stressed flanking vowels. There is a coarticulatory effect of stressed vowels that results in the shortening of following (and to a lesser extent preceding) unstressed transconsonantal vowels. Thus, stressed vowels exert a substantial amount of coarticulatory influence on unstressed vowels, especially on following unstressed vowels. There were no other cues in the familiarization stream to aid in the storage and recognition of syllable sequences.

Since it is not possible for a human talker to produce the entire familiarization string in one breath and still maintain the desired prosodic contour from beginning to end, the familiarization string was constructed from smaller units. All source items were naturally produced by a female native speaker of English and recorded on a Marantz cassette recorder with a close talking microphone. To create the naturally produced string, the larger familiarization sequence was divided into three syllable CVCVCV strings with medial word level stress (prominence of a syllable within the word). In order to preserve a natural sounding contour, each sequence was produced within a 5-syllable window which corresponded to the prosodic pattern of the English word 'unbeLIEvable' (stress is denoted by capital letters).

The crucial three syllable string was then cut out from the center of the 5 syllable frame at zero-crossings of the waveform and spliced together with the other items, yielding the familiarization string. For example, 'kugaBIgamu' was produced with stress on the syllable 'BI.' 'gaBIga' was then cut from the utterance. The sequence 'gamuNEpoku' was produced with stress on the 'NE' syllable and 'muNEpo' was cut from the string. The two sequences were spliced together to create the string 'gaBIgamuNEpo.' This process was used repeatedly for all nine tri-syllabic sequences in order to create the familiarization string. The familiarization string is in (1) where syllables in capital letters correspond to stress.

(1) BIgamuNEpokuTAnedoKUlepoGAdoneMUtaleDObitaPOmubiLEkuga

All infants were presented with the exact same, appropriately coarticulated, familiarization stimuli played in a two-minute continuous loop.

Following the familiarization phase, the infants were presented with test items. The infants were randomly assigned to one of two test groups: the group presented with *Appropriately Coarticulated* test sequences, or the group presented with *Miscoarticulated* test sequences. Both test groups were presented with the same two types of syllable sequences: three-syllable sequences that correspond to medially stressed items in the familiarization phase (*Repeated* test sequences), and three-syllable sequences that were created by concatenating syllables that never occurred adjacently in the familiarization strin, but did occur separately (*Novel* test sequences).
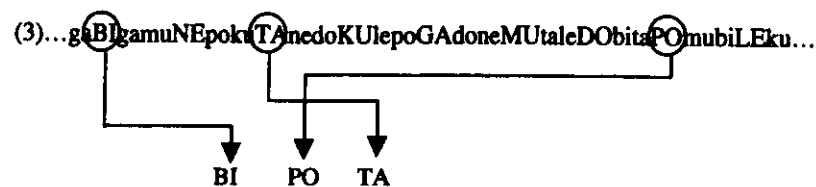
(2)	Repeated Test Sequences	Novel Test Sequences
	MUNEPO	BIPOTA
	LEDOBI	LEMUDO
	TAPOMU	GANEKU
	DOKULE	KUBINE

The Repeated sequences correspond to medially stressed sequences that occurred in the familiarization string, removing any left or right parsing biases. All items were relatively flat in their prosodic contour. How this was accomplished will now be discussed in our consideration of the construction of the test stimuli.

To construct the Appropriately Coarticulated test items, new recordings of the relevant syllables were used to create the repeated and novel sequences. Each syllable was produced in a correctly coarticulated frame: a CVC frame, in the case of the first two syllables, or a CV frame for the final syllable. For example, 'BIPOTA' was produced as 'BIP,' 'POT' and 'TA' syllables. This retained the coarticulation information of the consonants on the vowel. Each CV syllable was then cut from the CVC frame at the waveform zero crossing and spliced together with the following onset consonant. For example, the 'p' was cut from 'BIP,' and the 't' was cut from 'POT.' The remaining CV syllables were then all spliced together to create 'BIPOTA.' The rationale for splicing off the coda and then splicing the CV with the following onset consonant was to maintain the onset production of

the syllables, since there are no codas in the familiarization string. Further, since the items in the test phase were not presented with one syllable prominent relative to the other syllables in the string, it was necessary to produce monosyllabic utterances in order to maintain equal prominence across syllables. Lastly, in order to ensure proper duration of stop consonants in these test items, the stop closure length was evaluated based on each stop's second-syllable production in the familiarization recordings. That is, the closure length for the stops in the test sequences was determined by the pre-tonic occurrence of each stop in the familiarization stimuli. This technique was used to create all of the Repeated and Novel Sequences for the Appropriately Coarticulated Group.

The Miscoarticulated test sequences were created by producing a separate recording, using the same speaker, of the 5-syllable frames that matched the frames used in making the familiarization stream. To create the 3-syllable sequence 'taPOmu,' the 5-syllable frame 'bitaPOmubi' was recorded. Three-syllable sequences were then created by extracting syllables from these sequences and splicing them together. Specifically, the stressed syllable from the frame was cut out of the utterance at waveform zero-crossings. The Repeated and Novel test sequences were then created by splicing together three stressed syllables excised from the separate (unpresented) recording of the familiarization string. The creation of a new pseudo-familiarization string ensured that while the stimuli sounded similar to the presented familiarization string, they were not acoustic duplicates. The stressed instances of the syllables were used in order to retain full vowel quality and maintain a relatively flat prosodic contour. This way, the miscoarticulated sequences *and* the coarticulated sequences both had the same vowel qualities and prosodic contour. Additionally, by creating a new string, the items were crucially created with coarticulatory cues that, while appropriate for familiarization, were inappropriate for the context in which they occur during the test phase. For example, each of the underlined stressed syllables in (3) were spliced together to create the sequence 'BIPOTA.'

(3)...gaBIgamuNEpokuTAnedoKUlepoGAdoneMUtaleDObitaPOmubiLEku...

BI PO TA

Thus, 'BI' had the coarticulatory information for a preceding and following 'ga' which is inappropriate for the syllable sequence 'BIPOTA.' Both of the test sequences, Repeated and Novel, fail to contain appropriate contextual information. Thus it is possible that the infants' ability to recognize these sequences may be affected.

Subjects. 24 infants with a mean age of 7 1/2 months ($SD = 0.46$) were tested. The infants were randomly assigned to one of two test groups: (i) Appropriately Coarticulated, Repeated and Novel Sequences, and (ii) Miscoarticulated, Repeated and Novel Sequences. Both test groups were presented with the same appropriately coarticulated familiarization string.

Procedure. A preferential listening procedure adapted from Kemler Nelson, Jusczyk, Mandel, Myers, Turk & Gerken (1995) was used. The experiment was conducted in a sound-attenuated room with the experimenter situated outside the room observing the infant's looking behavior on a video monitor and coding the infant's looking behavior using a keyboard connected to a computer. The experimenter was unable to hear the stimuli played in the testing room but wore headphones nonetheless. The infant sat on the caregiver's lap. The caregiver listened to music over headphones in order to mask the stimuli.

A video camera and a central light were directly in front of the infant. There were lights on either side of the infant with loudspeakers directly above. During the familiarization phase of the experiment, the infants were presented with a 2-minute continuous sequence. The *same* familiarization string (appropriately coarticulated) was presented to both groups. The center light flashed to orient the child's gaze to the front of the room. Once the child was looking at the central light, the experimenter pressed a key that initiated the flashing of a randomly chosen side light. As the infant looked towards the flashing side light, the experimenter pressed the appropriate key to indicate a head-turn. The light flashed until the infant looked away for 2 consecutive seconds, at which point it was extinguished and the center light began flashing again. The familiarization loop was played continuously as

the procedure with the lights repeated. Thus, only the lights, not the familiarization stimuli, were contingent on the infant's head-turn behavior during this phase of the experiment (Mintz, 1996; Saffran, Aslin & Newport, 1996). This permitted an uninterrupted presentation of the familiarization loop.

Following the familiarization phase, a contingency phase occurred in order to provide a correlation between lights, sounds, and head-turns. The contingency phase consisted of four trials. A blinking light on the front wall began each trial. Once the infant fixated on the light, the experimenter initiated a trial. At this point the central light was extinguished, and one of the lights on the two side walls began to blink. Once the infant made a head-turn of minimally 30° towards a randomly chosen flashing side light, a tone sequence was played from the loudspeaker on that side. The sequence was repeatedly presented, with a 500ms interval between each presentation, until the infant's head-turn deviated from the corresponding light for two seconds. When the two-second look-away criterion was met, the side light was extinguished, and the central light began blinking to initiate another test trial. After completing the contingency phase, the test phase was presented using the same procedure. The computer recorded orientation times for each of the test trials.

Eight test items were presented per group. The test items corresponded to four medially stressed tri-syllabic sequences in the familiarization string (Repeated Test Sequences) and four control items (Novel Test Sequences), and were either Appropriately Coarticulated or Miscoarticulated depending on the group. The order of presentation of the test items was randomized.

2.2 Results

Mean orientation times to the two types of test items were calculated for each infant. In the Appropriately Coarticulated Group, nine of the twelve infants had significantly longer looking times for the Repeated sequences (mean rank = 7.11, n = 9 vs. mean rank = 4.67, n = 3) according to the Wilcoxon Matched-Pairs Signed-Ranks Test (2-tailed $p < .05$) (Figure 1). A paired-samples t-test (Repeated Test Sequences vs. Novel Test Sequences) found a significant effect of item type ($t(11) = -2.29$, $p < .05$). Infants listened longer to syllable sequences which occurred together in the familiarization

know that infants attend to syllable sequencing (Morgan & Saffran, 1995; Saffran, Aslin, & Newport, 1996; among others). These results confirm that the infants were able to recognize sequences of syllables from the familiarization exposure. However, a different pattern emerges for the Miscoarticulated Test Group.
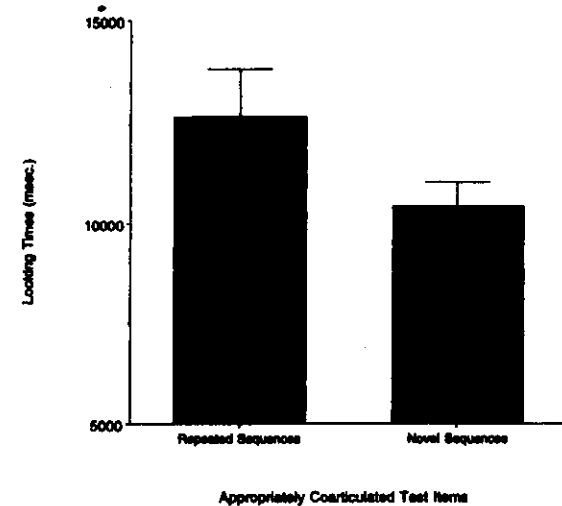

Figure 1: Appropriately Coarticulated Group's mean looking times for the Repeated and Novel Test Sequences.

Infants in the Miscoarticulated group showed no significant preference for either of the items (2-tailed $p = 1.0$; mean rank = 6.50) (Figure 2). According to a paired-samples t-test, there was no significant effect of item type ($t(11) = -.07$, $p = .944$). The test group presented with Miscoarticulated stimuli showed no preference for either the repeated or the novel sequences. To summarize, the test group that heard the Appropriately Coarticulated stimuli listened significantly longer to repeated sequences than to the novel sequences. However, the group that was presented with Miscoarticulated stimuli showed no preference for either the repeated or novel sequences.
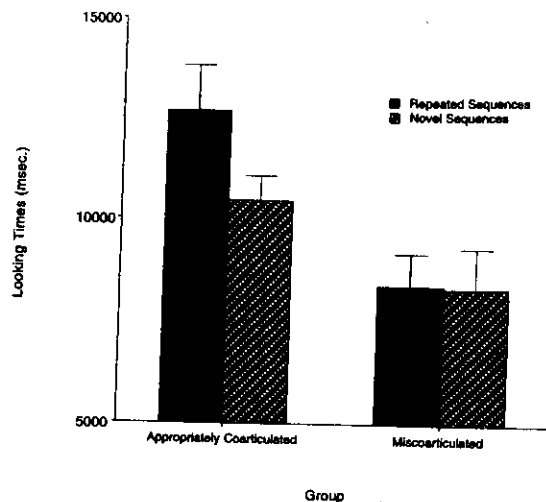
Figure 2: Mean looking times for both groups for the Repeated and Novel Test Sequences.

3.0 Discussion

The results of this experiment suggest that coarticulation is a salient aspect of the acoustic signal; that 7-month olds encode coarticulation information in representations of syllable sequences. Coarticulation seems to be a fundamental cue for sequence recognition. When coarticulatory cues are inappropriate, even with the sequential information is available, infants' recognition suffers.

There is, however, an alternative account of our data. It could be the case that infants are not processing the Miscoarticulated sequences as a signal comparable to that occurring during familiarization. In other words, they might not be processing the Miscoarticulated items as speech. However, research has shown that the sequential properties of non-speech stimuli are processed and responded to by infants of this age (Saffran, Johnson, Aslin & Newport, 1999).

Moreover, in a similar version of the coarticulation experiment with adult subjects, results indicate that segmental information is easily recoverable by adults presented with the Miscoarticulated test sequences. The adults were presented with the same familiarization sequence as the infants, played over headphones. Following the familiarization, they were

presented with either the Appropriately Coarticulated test sequences or the Miscoarticulated test sequences depending on the group they were randomly assigned to. They were asked to judge whether a sequence was familiar to them: 1 being familiar and 0 being not familiar.

The adults in both groups were able to recognize sequences which co-occurred during familiarization and gave higher familiarity ratings for these items over the novel sequences.
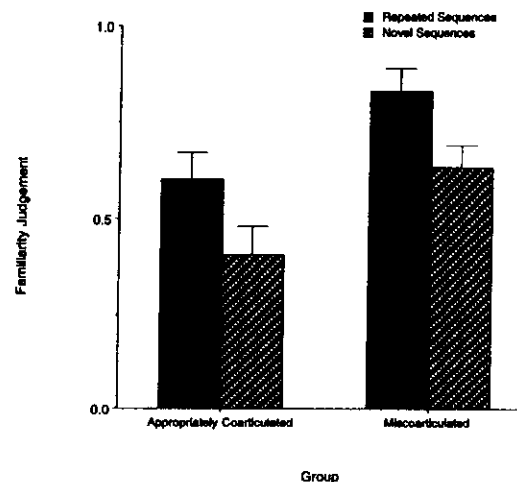

Figure 3: Adult familiarity judgments for the Repeated and Novel Test Sequences.

These results indicate that the repeated syllables are easily recognized by both groups of adults as familiar sequences, suggesting that the Miscoarticulated sequences are in fact processed as speech. Thus, the alternative explanation cannot account for both the adult and infant data.

The results suggest that infants are sensitive to specific sub-segmental properties of the acoustic signal that carry information about coarticulation. Moreover, we propose that the coarticulation information is processed and stored as part of the sequential information. Therefore, if the coarticulation information stored in the memory of the familiarization string does not match that of the test strings (as is the case for the Miscoarticulated Group), then repeated sequences do *not* cue recognition of the familiarization sequences.

There are further questions that arise as a result of this study. The data do not speak to the question of whether cues in the test stimuli must simply match the cues in the familiarization stream, or if they also must be appropriate for natural speech. For example, if infants were familiarized with a Miscoarticulated string, would the infants then prefer the Miscoarticulated repeated sequences? In other words, is it really the case that sequential memory and recognition for syllables hinges on speech being 'speech-like' in its coarticulatory patterning, or is it simply that the cues must *match* those of the familiarization string?

Finally, we can consider the question of whether coarticulation information is always important. It might be the case that coarticulation is weighted more heavily when segmentation is also happening, as we presume it is to some degree in the present experiment. One could test whether the group differences would still hold if infants were familiarized to a discrete (non-continuous) presentation of items, or whether the Miscoarticulation group would then also recognize the repeated sequences.

In conclusion, the results of this experiment demonstrate that infants' are sensitive to coarticulation information. When this information is available, infants store and use coarticulation cues for recognizing sequences. If the sequences do not have the appropriate coarticulatory cues, then infants' ability to recognize previously heard sequences diminishes. Thus, coarticulation information enhances infants' ability to recognize syllable sequences.

Endnote

References

Eimas, P. D., J. L. Miller, Jusczyk, P.W. (1987). On infant speech perception and the acquisition of language. *Categorical Perception: The Groundwork of Cognition*. S. Harnad. Cambridge, Cambridge University Press: 161-195.

Johnson, E. K.& Jusczyk, P.W. (in press) Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*.

Jusczyk, P.W. (1994). Infant speech perception and the development of the mental lexicon. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. J. Goodman and H. Nusbaum. Cambridge, MA, MIT Press: 227-270.

Jusczyk, P.W. (1997). *The Discovery of Spoken Language*. Cambridge, Mass.: MIT Press.

Jusczyk, P.W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, 64, 675-687.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., &

Gerken, L. (1995). The Headturn Preference Procedure in for testing auditory perception. *Infant Behavior and Development, 18, 111-116*.

McQueen, J.M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21-46.

Mintz, T. H. (1996). *The roles of linguistic input and innate mechanisms in children's acquisition of grammatical categories*. Unpublished Doctoral Dissertation, University of Rochester.

Morgan, J.L. & Saffran, J.R. (1995). Emerging Integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, 66, 911-936.

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

Saffran, J. R., Johnson, E. K., Aslin, R.N., & Newport, E.L. Statistical learning of tone sequences by human infants and adults. *Cognition, 70* (1) (1999) pp. 27-52

Werker, J. F. and J. Pegg (1992). Infant speech perception and phonological acquisition. *Phonological Development: Models, Research, Implications*. C. A. Ferguson, L. Menn and C. Stoel-Gammon. Timonium, MD, York Press: 131-164.