INFANCY, *7*(1), 99–109

English-Learning Infants' Representations of Word Forms With Iambic Stress

Elizabeth K. Johnson
Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands

Retaining detailed representations of unstressed syllables is a logical prerequisite for infants' use of probabilistic phonotactics to segment iambic words from fluent speech. The head-turn preference study was used to investigate the nature of English-learners' representations of iambic word onsets. Fifty-four 10.5-month-olds were familiarized to passages containing the nonsense iambic word forms *ginome* and *tupong*. Following familiarization, infants were either tested on familiar (*ginome* and *tupong*) or near-familiar (*pinome* and *bupong*) versus unfamiliar (*kidar* and *mafoos*) words. Infants in the familiar test group (familiar vs. unfamiliar) oriented significantly longer to familiar than unfamiliar test items, whereas infants in the near-familiar test group (near-familiar vs. unfamiliar) oriented equally long to near-familiar and unfamiliar test items. Our results provide evidence that infants retain fairly detailed representations of unstressed syllables and therefore support the hypothesis that infants use phonotactic cues to find words in fluent speech.

Contrary to common intuition, the boundaries of spoken words are not consistently and reliably marked by pauses or any other acoustic cues (Cole & Jakimik, 1980). The physical nature of word boundaries stands in stark contrast to our perception of word boundaries. Indeed, hearing words is so effortless that it may seem as if speakers place tiny pauses between words. However, this illusion is easily broken by listening to an unfamiliar language. Conversations in unfamiliar languages sound very fast, and determining where one word ends and the next begins is virtually impossible. Listening to foreign languages helps us understand the continuous nature of speech because hearing words in our native

Requests for reprints should be sent to Elizabeth K. Johnson, Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. E-mail: elizabeth.johnson@mpi.nl

language is dependent on our sensitivity to a wide array of language-specific cues to word boundaries.

Lexical stress is one important language-specific cue to word boundaries. In English, for example, the majority of content words carry word-initial stress (e.g., *doggie;* Cutler & Carter, 1987), and adult English speakers are biased to perceive stressed syllables as word onsets (Cutler & Butterfield, 1992). This bias serves as a useful segmentation heuristic. The emergence of word segmentation abilities in English-learning infants has been linked to infants' developing sensitivity to the predominant word-initial stress pattern of English. Studies have shown that infants begin segmenting words by 7.5 months of age (Jusczyk & Aslin, 1995), at about the same age they first demonstrate a listening preference for trochaic (stress-initial bisyllables) over iambic (stress-final bisyllables) words (Jusczyk, Cutler, & Redanz, 1993). Additional evidence for the importance of stress cues in early word segmentation comes from young infants' segmentation errors. Whereas 7.5-month-olds reliably extract trochaic words (e.g., *doctor*) from fluent speech, they systematically fail to segment iambic words (e.g., *guitar;* Jusczyk, Houston, & Newsome, 1999). At the same time, they mistake trochaic cross-word sequences for words. For example, if 7.5-month-olds are familiarized with passages containing the word *guitar* consistently followed by the word *is* (e.g., …*guitar is in* … *guitar is mine* … *guitar is too*), they perceive the trochaic cross-word sequence *taris* as a word.

Although stress cues provide a useful segmentation heuristic in English, English also has a sizable minority of words that do not carry word-initial stress (e.g., *guitar*). Thus, infants must eventually overcome their reliance on stress cues to determine word boundaries. Jusczyk et al. (1999) investigated this issue by testing 10.5-month-olds on the same materials they had used with 7.5-month-olds. They found that 10.5-month-olds segment both iambic and trochaic words equally readily. Moreover, Jusczyk et al. found that 10.5-month-olds no longer mistake reoccurring trochaic cross-word sequences (e.g., *taris*) for words. In short, Jusczyk et al. found that English learners overcome their overreliance on stress cues by 10.5 months.

The discovery that 10.5-month-olds can reliably segment iambic words from fluent speech naturally led researchers to wonder how infants accomplish this task. One explanation offered by Jusczyk et al. (1999) is the use of probabilistic phonotactics, or constraints on the likelihood of phoneme sequences occurring across versus within word boundaries. Research has shown that adult English speakers use probabilistic phonotactics to spot words in speech (McQueen, 1998; Vitevitch & Luce, 1999). Support for the hypothesis that infants use probabilistic phonotactics to segment iambic words from fluent speech comes from studies using the head-turn preference procedure (HPP). Mattys and Jusczyk (2001) familiarized 9-month-olds to one passage containing a monosyllabic nonsense word flanked by strong phonotactic cues (consonant clusters typically found between, as opposed to within, word boundaries in a corpus of child-directed speech, such as [vt] and [fh]), and one passage containing a monosyllabic nonsense word flanked by weak phonotactic cues (consonant clusters typically found within, as opposed to between, word boundaries in a corpus of child-directed speech, such as [ft] and [vn]). The experiment was designed so that the same nonsense word (e.g., *tove*) could be flanked by either strong phonotactic cues (e.g., *brave tove trusts*) or weak phonotactic cues (e.g., *gruff tove knows*). Mattys and Jusczyk found that infants segmented the nonsense words when they were flanked by strong phonotactic cues but not when they were flanked by weak phonotactic cues. Thus, they argued that 9-month-olds' sensitivity to probabilistic phonotactics could play an important role in enabling infants to eventually overcome their overreliance on stress cues.

Given that infants become sensitive to probabilistic phonotactics at 9 months of age (Jusczyk, Luce, & Charles-Luce, 1994), just prior to the ability to segment iambic words, it seems reasonable to argue that probabilistic phonotactics play an important role in 10.5-month-olds' ability to segment iambic words from running speech. However, the feasibility of this argument rests on a crucial assumption, namely, that infants extract fairly detailed representations of unstressed syllables. Using probabilistic phonotactics to segment iambic words would be impossible if infants' representations of unstressed syllables were not adequately specified. For example, [vt] is a sequence that occurs more often across, as opposed to within, word boundaries, whereas [vn] is a sequence that occurs more often within, as opposed to across, word boundaries. Imagine that a 10-month-old hears the phrase *arrive today.* The occurrence of the cross-word sequence [vt] could help an infant place a word boundary between *arrive* and *today.* However, if the initial [t] of *today* is not represented in adequate detail to distinguish it from another segment such as an [n], then phonotactic cues would be of little help in finding the onset of *today* because the sequence [vn] is a within-word sequence. In short, if infants do not represent unstressed syllables in sufficient detail (e.g., detect single-segment alterations to the onsets of iambic words), then phonotactic knowledge would be of little help in extracting iambic words from fluent speech.

The existing literature does not contain any studies specifically designed to ask whether English-learning infants represent unstressed syllables in sufficient detail to detect single-segment alterations to the initial syllable of iambic words. However, the current literature does contain many findings that are relevant to this question. On the one hand, evidence suggests that unstressed syllables are not represented in great detail. Young children frequently omit function words and other unstressed syllables from their productions (Echols & Newport, 1992), and the acoustic characteristics of stressed syllables are more pronounced than those of unstressed syllables (Lehiste, 1970). Moreover, French-learning 11-month-olds reportedly possess less than detailed representations of familiar iambic words

(Halle & Boysson-Bardies, 1996). On the other hand, toddlers seem to perceive and process function words and other unstressed syllables before they begin producing them (Fernald & Zangl, 2003; Gerken & McIntosh, 1993; Gerken, Remez, & Landau, 1990; Kirk & Seidl, 2004). There is also evidence that infants retain fairly detailed representations of both stressed and unstressed syllables. For example, infants familiarized with passages containing the word *cup* do not subsequently recognize either *tup* or *cut* as familiar (Jusczyk & Aslin, 1995; Tincoff & Jusczyk, 1996). And 10.5-month-olds perceive iambic words as whole units rather than fragments (e.g., *tar* is not segmented from passages containing repetitions of *guitar*). Moreover, both English- and German-learning infants appear to represent highly frequent unstressed function words in at least some detail. Shi, Werker, and Cutler (2003) found that 13-month-olds listen longer to isolated repetitions of a content word when it is preceded by a real functor (e.g., *the*) rather than a prosodically similar nonce functor (e.g., *kuh*). Hoehle and Weissenborn (2003) found that German-learning 8-month-olds familiarized to isolated tokens of *bis* and *sein* subsequently preferred to listen to passages containing *bis* and *sein* over passages containing *von* and *das.*

Taken together, these findings demonstrate that infants not only represent stressed syllables in great detail, but they also perceive unstressed syllables as more than phonetically unspecified weak beats in the speech stream. At the same time, these findings do not necessarily indicate that infants represent the unstressed onsets of newly learned iambic words in great detail. Indeed, the possibility that 10.5-month-olds might accept *puitar* as an acceptable token of *guitar* is still perfectly consistent with the previously mentioned evidence. If infants do not represent unstressed syllables in adequate detail to distinguish between *guitar* and *puitar,* then phonotactics would be of limited use in finding iambic words in speech. Thus, this study is specifically designed to address this question: Can English-learning infants extract sufficiently detailed representations of unstressed syllables to detect single-segment changes to a newly learned iambic word?

In this study, the HPP was used to familiarize 10.5-month-olds to two passages, each containing six repetitions of a nonsense iambic word (*ginome* and *tupong*). Half of the infants were assigned to the familiar test group, and half were assigned to the near-familiar test group. Infants in the former group were tested on two familiar (*ginome* and *tupong*) and two unfamiliar (*kidar* and *mafoos*) words. Infants in the latter group were tested on two near-familiar words that differ from the familiar words by a single segment (*pinome* and *bupong*) as well as two unfamiliar words (*kidar* and *mafoos*).

As in Jusczyk et al. (1999), we expected infants to orient longer to familiar than unfamiliar test items. Thus, in this study, infants in the familiar test group should orient longer to familiar than unfamiliar test items. In the near-familiar test group, however, there are two plausible outcomes. If infants' representations of unstressed syllables are not detailed, then they may mistake the near-familiar items for familiar items, resulting in longer orientation times to near-familiar than unfamiliar test items. On the other hand, if infants' representations of unstressed syllables are sufficiently detailed to detect single-segment alterations, then they will perceive the near-familiar (e.g., *pinome*) and familiar (e.g., *ginome*) items as two different words. In this case, past research suggests that they will demonstrate no orientation preference for near-familiar over unfamiliar test items (Jusczyk & Aslin, 1995; Tincoff & Jusczyk, 1996). The latter outcome would suggest that infants extract detailed representations of unstressed syllables, which in turn would lend strength to the argument that probabilistic phonotactics are a useful tool in learning to segment iambic words from fluent speech.

METHOD

Participants

Fifty-four American-English-learning 10.5-month-olds (range = 10–11 months) from the Baltimore–Annapolis region were randomly assigned to either the familiar or near-familiar test group. The 27 infants in the familiar test group (15 girls) had a mean age of 318 days (range = 304–333 days). The 27 infants in the near-familiar test group (11 girls) had a mean age of 317 days (range = 304–335 days). The data from 9 additional infants were excluded for the following reasons: fussiness (8) and average orientation times less than 3 sec (1). Parental consent was obtained for all participants.

Stimuli

A female speaker who was naive to the purpose of the study recorded both passages in an infant-directed manner. Each six-sentence passage contained one target word per sentence, each flanked by strong phonotactic cues (see the Appendix). The *ginome* and *tupong* passages were 18.9 and 18.2 sec long, respectively. Acoustic analyses were carried out to confirm the target words' iambic stress pattern (see Table 1; due to measurement difficulties, the [r] in *kidar* could not be separated from the preceding vowel and was therefore included in the vowel measurements). The second syllables were marked by longer duration and higher amplitude. Unexpectedly, the second syllables were not systematically higher in pitch than the first syllables. However, this is explainable given the unreliable nature of fundamental frequency (F0) as a cue to stress in running speech (Lehiste, 1970). The speaker also recorded approximately 20 isolated repetitions of each of the iambic test items: *ginome* [gɪnom], *pinome* [pɪnom], *tupong* [tupɑŋ], *bupong* [bupɑŋ], *kidar* [kɪdɑr], and *mafoos* [mɑfus]. Fifteen

TABLE 1
Mean Acoustic Values for Vowels in Target Words
Produced in Familiarization Passages

	Duration (msec)				*Amplitude (dB)*				*Pitch (Hz)*			
	Syllable 1		*Syllable 2*		*Syllable 1*		*Syllable 2*		*Syllable 1*		*Syllable 2*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Ginome	45	9	167	38	68	2	72	2	314	54	304	99
Tupong	39	9	178	7	65	1	71	2	265	33	256	41

TABLE 2
Mean Acoustic Values for Vowels in Target Words Produced in Test Lists

	Duration (msec)				*Amplitude (dB)*				*Pitch (Hz)*			
	Syllable 1		*Syllable 2*		*Syllable 1*		*Syllable 2*		*Syllable 1*		*Syllable 2*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Ginome	56	11	349	28	73	3	78	2	285	69	398	109
Tupong	40	10	327	86	68	5	77	1	279	47	364	97
Pinome	48	14	324	25	72	5	77	3	329	102	406	115
Bupong	45	12	352	36	68	5	77	1	299	69	402	115
Kidar	36	10	474	42	67	4	76	2	320	76	378	56
Mafoos	56	11	198	24	70	3	79	2	319	85	497	75

clearly articulated tokens of each were chosen for use in the study. Six test lists, each consisting of 15 different tokens of the same word, were created. Care was taken to ensure that the test lists were acoustically similar and produced with iambic stress (see Table 2). Test lists were on average 16 sec long (range = 15.7–16.2 sec).

All recordings were made using a Shure microphone in a sound-shielded booth. They were digitized on a CSL 150 workstation at a 20 kHz sampling rate via a 16-bit analog-to-digital converter. The recordings were digitally transferred to a Macintosh Quadra 650 computer for playback during the experiment. The recordings were played at a comfortable listening level (approximately 72 dB SPL, according to a Quest Model 215 sound meter). The computer controlled the presentation of the sound files during the experiment. A 16-bit analog-to-digital converter was used to re-create the audio signal at a 20 kHz sampling rate. The output was fed through antialiasing filters and a Kenwood audio amplifier (KA 5700) to the two 7 in. (18 cm) Cambridge Soundworks loudspeakers mounted on the side walls of the test booth.

Procedure

Infants were tested using the same version of the HPP first used by Jusczyk and Aslin (1995). In this procedure, the infant sits on a caregiver's lap in the center of a three-sided booth. Three lights are mounted at eye level: a green light on the front panel and a red light on each of the two side panels. Speakers are located behind the red lights. The experimenter observes the infant through a tiny peephole and relays the infant's looking behavior to the computer via a button box. Each trial begins with the center green light flashing. When the infant orients to the center light, it stops flashing and one of the side lights immediately begins flashing. Once the infant orients at least 30° toward the flashing light, a sound file is presented from the speaker hidden behind the flashing light. The sound file continues to play until either the infant looks away for more than 2 sec or the sound file ends. The dependent measure in this procedure is orientation time toward test items.

The experiment consisted of two phases: familiarization and test. During familiarization, the *ginome* and *tupong* passages played alternately until the infant accrued 45 sec of orientation time toward each passage. Twelve test trials were presented during the test phase (three trials for each of the four test items). Test trials were blocked and presented in random order within those three blocks. For infants assigned to the familiar test group, half of the trials consisted of repetitions (up to 15 per trial) of two familiar words (*ginome* and *tupong*); the other half consisted of comparable repetitions of two unfamiliar words (*kidar* and *mafoos*). For infants assigned to the near-familiar test group, half of the trials consisted of repetitions (up to 15 per trial) of two near-familiar words (*pinome* and *bupong*); the other half consisted of comparable repetitions of two unfamiliar words (*kidar* and *mafoos*).

Both the experimenter and the caregiver wore well-insulated tight-fitting headphones (Peltor Aviation Headset 7050) over which loud masking music was played so they could not tell which stimulus was playing at any given time (see Kemler Nelson et al., 1995, for data on the efficacy of this masking procedure).

RESULTS

Mean orientation times to familiar (or near-familiar) and unfamiliar test items were calculated for each of the 54 participants (see Figure 1). On average, infants in the familiar test group oriented to the familiar words for 8.51 sec ($SD = 2.2$) and to the unfamiliar words for 7.3 sec ($SD = 2.6$). In contrast, infants in the near-familiar test group oriented to the near-familiar words for 6.6 sec ($SD = 2.2$) and to the unfamiliar words for 7.3 sec ($SD = 2.5$). A 2 (test group: familiar vs. near-familiar) × 2 (test item type: familiar or near-familiar vs. unfamiliar) mixed design analysis of variance (ANOVA) revealed no significant effect of test item type, $F(1, 52) < 1$. In addition, there was no significant effect of test group, $F(1, 52) = 2.76$, $p > .10$.

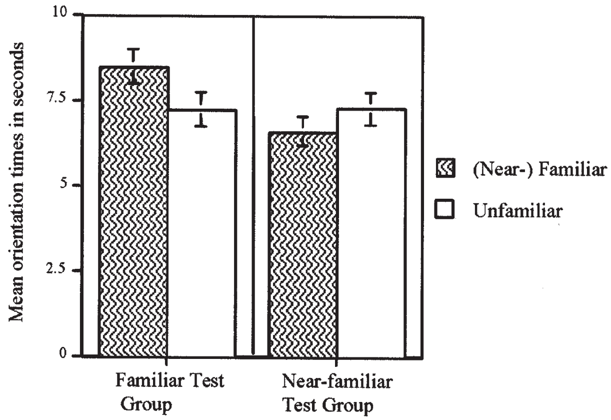

FIGURE 1 Mean orientation times in seconds broken down by test item type (familiar or near-familiar vs. unfamiliar) and test group (familiar test group vs. near-familiar test group). Error bars indicate standard error.

Most important, however, there was a significant Test Item Type × Test Group interaction, $F(1, 52) = 8.2, p < .01$. Planned comparisons revealed a significant effect of test item type in the familiar test group, $F(1, 26) = 5.6, p < .05$. There was no effect of test item type in the near-familiar test group, $F(1, 26) = 2.6, p > .10$.

These results indicate that infants in the familiar test group oriented longer to familiar test items than unfamiliar test items. However, infants in the near-familiar test group failed to orient longer to near-familiar test items than unfamiliar test items. To ensure that these effects were not driven by unequal familiarization times to the two familiarization passages, a 2 (test group: familiar vs. near-familiar) × 2 (familiarization passage: *ginome* vs. *tupong*) mixed design ANOVA was carried out. There was no effect of familiarization passage; that is, there was no significant difference in infants' orientation to the *ginome* ($M = 54.7$ sec) versus *tupong* ($M = 54.0$ sec) passages, $F(1, 52) < 1$. There was no interaction between test group and familiarization passage, $F(1, 52) < 1$.

DISCUSSION

The goal of this study was to determine whether English-learning 10.5-month-olds extract detailed representations of novel iambic words from fluent speech. Infants tested in the familiar test group listened longer to familiar iambic words than unfamiliar words, replicating the results of Jusczyk et al. (1999). Infants in the near-familiar test group, on the other hand, oriented equally long toward all test items. In other words, they did not false alarm to near-familiar words. Thus, infants' representations of unstressed syllables appear to be adequately detailed to detect single-segment alterations (e.g., *ginome* vs. *pinome*). This finding adds strength to the argument that 10.5-month-olds use probabilistic phonotactics to begin segmenting iambic words from fluent speech. Note, however, that both of the near-familiar test items used in this study contained word-initial single-segment alterations involving changes in two features (voice and place). Further research will be needed to determine if infants' representations of unstressed syllables are detailed enough to detect single-segment as well as single-feature alterations in both word-initial and word-final position (e.g., *ginome* vs. *kinome* as well as *ginome* vs. *ginone*). This latter comparison is particularly crucial given evidence that infants are more sensitive to word-final versus word-initial segments (Jusczyk, Goodman, & Bauman, 1999; Zamuner, 2004). It may also be important to consider whether all feature changes are equally salient to infants (e.g., voice vs. place). By further exploring how infants represent unstressed syllables, we will begin to further understand the role of phonotactic cues in early word segmentation.

ACKNOWLEDGMENTS

This investigation was supported by National Institute of Child Health and Human Development Research Grant 15795 and by National Institute of Mental Health Senior Scientist Award 01490 to Peter W. Jusczyk. Portions of this research were presented at the 13th biennial meeting of the International Society on Infant Studies, Toronto, Ontario, Canada, April 2002. I thank Peter W. Jusczyk for inspiring this project, Ann Marie Jusczyk and Courtney VanCleefe for help in running and recruiting participants, and Joanne L. Miller and Tania Zamuner for many insightful comments made on previous drafts of this article. All errors are, of course, my own.

REFERENCES

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 136–163). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31,* 218–236.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133–142.

Echols, C., & Newport, E. (1992). The role of stress and position in determining first words. *Language Acquisition, 2,* 189–220.

Fernald, A., & Zangl, R. (2003, November). *Sensitivity to function word morphemes in online sentence processing: Developmental changes from 18 to 36 months.* Talk presented at the Boston University Conference on Language Development, Boston.

Gerken, L., & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology, 29,* 448–457.

Gerken, L., Remez, R. E., & Landau, B. (1990). Function morphemes in young children's speech-perception and production. *Developmental Psychology, 26,* 204–216.

Halle, P., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicons. *Infant Behavior and Development, 19,* 463–481.

Hoehle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science, 6,* 122–127.

Jusczyk, P. W., & Aslin, R. N. (1995). Infant's detection of sound patterns of words in fluent speech. *Cognitive Psychology, 29,* 1–23.

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

Jusczyk, P. W., Goodman, M. B., & Bauman, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language, 40,* 62–82.

Jusczyk, P. W., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39,* 159–207.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33,* 630–645.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The headturn preference procedure for testing auditory perception. *Infant Behavior and Development, 18,* 111–116.

Kirk, C., & Seidl, A. (2004). Perception and production of unstressed syllables: Implications for lexical representations. In *Proceedings of the 28th Annual Boston University Conference on Language Development* (pp. 318–327). Boston: Cascadilla Press.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78,* 91–121.

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

Shi, R., Werker, J., & Cutler, A. (2003, August). *Function words in early speech perception*. Talk presented at the 15th International Conference of Phonetic Sciences, Barcelona, Spain.

Tincoff, R. L., & Jusczyk, P. W. (1996, July). *Are word-final sounds salient for infants?* Poster presented at the Fifth Conference on Laboratory Phonology, Evanston, IL.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood lexical inhibition in spoken word recognition. *Journal of Memory and Language, 40,* 374–408.

Zamuner, T. S. (2004, May). *Infants' perception of place of articulation and voicing in word- initial and word-final position.* Poster presented at the International Conference on Infant Studies, Chicago.

APPENDIX

The old stone ginome could tell a great story.
One ginome clowned around with a spinning bear.
The pine ginome carelessly tangoed all through the night.
We can see how well the short tune ginome cooks.

Some lone ginome cradled a squirrel in his arms.
The tall tin ginome cut flowers for his cousin.

Live tupong teach painting every other night.
I plan to cartwheel with the mauve tupong troop.
We like the green weave tupong tie around their forts.
The brave tupong told us singing is an art form.
Some suave tupong traded a soccer ball for us.
The love tupong tap-danced at the talent show.