# Exploring statistical learning by 8-month-olds: The role of complexity and variation

Elizabeth K. Johnson and Peter W. Jusczyk*

*Introduction*

Recent studies suggest that young infants possess powerful statistical learning mechanisms that they can use to find word boundaries in fluent speech. For example, in a landmark study by Saffran, Aslin, and Newport (1996), the Headturn Preference Procedure was used to familiarize English-learning 8-month-olds to an artificial language containing no cues to word boundaries other than transitional probabilities between syllables. The language they used contained four trisyllabic words that were created by combining twelve CV syllables. The four trisyllabic words were concatenated into a continuous two-minute stream of speech containing no pauses between words (e.g.….pabikudaropitibudopabikugolatu….). This speech stream was intended to mimic a real language learning environment, since real speech does not contain reliably placed pauses between words either (see Jusczyk, 1997, for a thorough review of the word segmentation problem). In Saffran et al.'s stimuli no single word repeated itself in immediate succession, and no syllable was used in more than one word. Therefore, the transitional probability between two syllables within a word was equal to 1.0, whereas the transitional probability between two syllables crossing a word boundary was equal to .33 (if *pabiku* and *daropi* are words, then the transitional probability between *bi* and *ku* is 1.0, whereas the transitional probability between *ku* and *da* is .33). Saffran et al. (1996) found that after familiarization with this language, infants listened significantly longer to statistical part words (e.g. *kudaro* if *pabiku* and *daropi* are statistical words) than statistical words. Saffran et al. interpreted this novelty effect as evidence that infants are very good at statistical learning. In addition, they argued that statistical learning might play an important role in early word segmentation. Additional work by the same group of authors demonstrated that infants are also able to track conditional probabilities between syllables (Aslin, Saffran, and Newport, 1998).

The finding that infants can segment an artifiical language containing no cues to word boundaries other than the strength of the transitional probabilities between syllables is an impressive finding. However, the language used by Saffran et al. (1996) was highly simplified with respect to natural language input. Thus, one cannot automatically assume that infants can used statistical cues to segment a natural language, especially given past studies demonstrating that statistical learning effects disappear rapidly as the complexity of stimulus patterns increase (Bruner, 1973). Therefore, in the following two experiments, we ask whether infants' statistical learning abilities are robust enough to deal with a slightly more complex artificial language. The artificial language we use in the following two experiments will be very similar to that used by Saffran et al. (1996). The only difference is that the language used in this study, like real language, will contain words of variable length. Thus, instead of being familiarized to an artificial language containing four trisyllabic words, the infants in our study are familiarized with an

* These data were compiled and written by Elizabeth K. Johnson (zab@jhu.edu) following the death of PWJ.

artificial language containing four words of variable length (1 two-syllable word, 1 four-syllable word, and 2 three-syllable words). If infants' ability to track transitional probabilities between syllables is robust enough to deal with natural language, than one would expect that infants would be able to segment our slightly modified version of Saffran et al.'s language.

## Experiment 1

In Experiment 1, infants are familiarized with a modified version of the artificial language used by Saffran et al. (1996). The language used in this experiment is nearly identical to that used by Saffran et al., however, the language used in the current study contains words of variable length. The transitional probabilities defining within-word versus cross-word syllable sequences is identical to those in Saffran et al. (1996). As in Saffran et al., we use the Headturn Preference Procedure to familiarize infants to a 2.5-minute continuous stream of speech. Following familiarization, infants listening preferences to words versus part words will be tested. If infants' ability to track transitional probabilities is truly robust enough to handle the variation seen in natural language, we predict that infants will display a novelty effect for part words during the test phase. If however, the results of Saffran et al. were highly dependent on the simplistic structure of their artificial language, then we predict that infants may have no preference for part words over statistical words.

### Methods

*Participants*

Twenty-four English-learning 8-month-olds from the Baltimore-Annapolis region were tested (11 Males and 13 females). The infants were approximately 8.0 months old (range 7.5-8.5), with a mean age of 248 days (Range: 233-260; SD=6.6). The data from an additional 3 infants were excluded due to fussiness. Participants were recruited to come to the lab via letters mailed by the Maryland State Health Department, advertisements in local parenting magazines, and referrals from past participants. Parental consent was obtained for all participants. In appreciation for their participation, all infants were given a small toy or certificate and a $7 travel reimbursement.

*Stimuli*

This study used the same 12 syllables used by Johnson and Jusczyk (2001). Two sets of four words each were created by stringing together these syllables. The first set of four words were used to form Artificial Language A: *pabiku*, *tibudo*, *gola*, and *tudaropi*. The second set was used to form Artificial Language B: *pigola*, *tudaro*, *biku*, and *tibudopa*. Note that in contrast to the language used by Saffran et al. (1996), each language used in the current study contained words of variable length: one 4-syllable word, one 2-syllable word, and two 3-syllable words.

A 2.5-minute speech stream was created for each language by concatenating together the four words in the language. No syllable was used in more than one word, and words were randomly concatenated together with the stipulation that no word occurred twice in succession. Therefore, the statistical cues to word boundaries were the same as those both in Saffran et al. (1996) and in Johnson and Jusczyk (2001). The transitional probability between all syllables within words was always equal to 1.0, and the transitional probability between any two syllables spanning the boundary between two words was always equal to .33. No pauses occurred between the words. Orthographically, one of the resulting speech streams would be represented as follows: pabikutibudopabikugolatudaropi….

Infants familiarized with Language A were tested on the following four test items: *gola*, *pabiku*, *dotu*, and *pitibu*. Infants familiarized with Language B were tested on a second set of four test items: *biku*, *tudaro*, *lati*, and *papigo*. Test items were concatenations of the individual syllables recorded for use in Johnson and Jusczyk (2001). Statistical part words were formed by taking the last syllable of one statistical word and combining it with one or more syllables from the beginning of another statistical word. All infants were tested on two statistical words and two statistical part words.

*Design*

All infants were randomly assigned to hear one of the two streams of speech (Language A or Language B). Twelve test trials were presented immediately after the familiarization phase. Half of the trials consisted of repetitions of a string of syllables corresponding to a statistical word (*pabiku* if familiarized with Language A; *tudaro* if familiarized with Language B); the other half consisted of repetitions of a string of syllables corresponding to a statistical partword (*pitibu* from *tudaropi* and *tibudo* if familiarized with Language A; *papigo* from *tibudopa* and *pigola* if familiarized with Language B). Note that the transitional probabilities between statistical part words and statistical words in this study were identical to those between the statistical part words and words used in the studies carried out by both Saffran et al. (1996) and Johnson and Jusczyk (2001). Therefore, if infants fail to segment the speech stream used in the current study, then it could not be due to the strength of the statistical cues defining word boundaries. Rather, it seems likely that by making the language slightly more complex by varying the length of the words used in the artificial language, the task of segmentation based purely on statistical cues will have become too difficult for infants. If simply varying the length of the words in the artificial language makes it impossible for infants to segment the language, then this does not bode well for infants' heavy reliance on transitional probabilities to segment natural speech.

*Procedure and Apparatus*

Infants were tested using the same version of the Headturn Preference procedure used by Saffran et al. (1996). The exact same apparatus used by Johnson and Jusczyk (2001) was used in the current study.

Infants sat in the center of a caregiver's lap. The caregiver was seated on a chair in the center of a three-sided booth made of white pegboard. A red light and a speaker were mounted at eye level on the center of each side panel, and a green light was located at eye level on the center of the front panel. During the familiarization phase, the green light flashed at the start of each trial. Once the infant oriented towards the green light, one of the two speech streams played from the two side speakers continuously until the end of the sound file was reached (approximately two and a half minutes). Throughout the familiarization, in order to entertain the infants, all three lights in the testing booth were lit and extinguished in response to the infants' looking behavior (see Johnson & Jusczyk, 2001, for additional methodological details). The test phase immediately followed the familiarization phase. Each of the 12 test trials (three trials for each of the four test items) began with the blinking center light. Once the infant oriented toward the green light, the green light stopped blinking and one of the two side red lights began blinking. Once the infant oriented toward the blinking light, a test item was repeated with a 500 ms ISI until the infant looked away from the blinking light for more than 2 consecutive seconds or until 15 repetitions of the test item had occurred. Thus, the infant essentially controlled how long he or she heard the test items. Test trials were blocked and presented in random order within those three blocks.

The experimenter recorded the direction and duration of infants' orientation via a botton box connected to a Macintosh computer. Computer software was responsible for the selection and randomization of the stimuli and for the termination of test trials. Both the caregiver and the experimenter wore tight-fitting headphones over which loud masking music was played (for a more thorough methodological discussion of the Headturn Preference Procedure, see Kemler Nelson et al., 1995). The dependent measure in this study was orientation time to different types of stimuli. As in Saffran et al. (1996) and Johnson and Jusczyk (2001), we predicted that infants' ability to segment the artificial language would be signaled by longer listening times to novel-sounding stimuli (e.g. part words).

All recordings were played at a comfortable listening level (approximately 72 dB SPL, according to a Quest (Model 215) sound meter. The audio output was generated from the digitized waveforms of the samples. A 16-bit D/A converter was used to recreate the audio signal. The output was fed through anti-aliasing filters and a Kenwood audio amplifier (KA 5700) to the two 7-inch Cambridge Soundworks loudspeakers mounted on the side walls of the testing booth.

<u>Results and Discussion</u>

Mean orientation times to the types of test items (statistical part words versus statistical words) during the test phase were calculated for each infant. Thirteen out of 24 infants had longer orientation times to statistical part words than statistical words. Mean orientation times broken down by word length and test item type were as follows: trisyllabic word ($M=7.18$ seconds; $SD=2.4$), bisyllabic word ($M=7.36$; $SD=2.4$), trisyllabic part word ($M=7.44$; $SD=2.9$), bisyllabic part word ($M=7.28$; $SD=2.7$). A 2 X 2 mixed design ANOVA with Condition (familiarized with Language A versus Language B) and test item (statistical part word versus statistical word) revealed no significant effect of test item, $F(1, 22) < 1$. There was a significant effect of group, $F(1, 22) = 5.06$, $p < .05$. However, importantly, there was no significant interaction between group and test item, $F(1,22) < 1$. Thus, it appears that English-learning 8-month-olds failed to segment an artificial language containing words of varying length.

## Experiment 2

In Experiment 1, English-learning 8-month-olds failed to segment an artificial language containing words of varying length. One possible explanation for this finding is that infants' ability to use statistical cues to segment speech is not robust enough to handle the variation seen in natural languages (e.g. words of varying length). However, another possible explanation is that the design of the study, rather than the complexity of the stimuli, made infants fail to segment the artificial language made up of words of different lengths. In all published segmentation studies that we are aware of, all test items contain the same number of syllables. Therefore, the fact that our test items varied in length could have been a possible confound in Experiment 1.[1] Thus, the results of Experiment 1 could be attributable to a quirk of the Headturn Preference Procedure rather than lack of the robustness of statistical learning in infants.

In Experiment 2, infants were tested using the same familiarization languages used in Experiment 1. However, instead of testing the 8-month-olds on four test items of variable length, we tested infants on four trisyllabic words (this was possible because each language contained two trisyllabic words). Thus, the only difference between this experiment and Experiment 1 in Johnson and Jusczyk (2001) is the variable length of the words in the artificial language. If the added complexity of variable word length does not hinder infants' ability to learn the statistical structure of the language, then infants should listen longer to novel part words (e.g. pigola from tudaropi and gola if familiarized with Language A; pabiku from tibudopa and biku if familiarized with Langauge B). On the other hand, if the results of Experiment 1 are not simply due to the variable length of the test items, then we should once again see no significant preference for novel part words over familiarized words during the test phase.

### Methods

*Participants*

Twenty-four English-learning 8-month-olds from the Baltimore-Annapolis region were tested (11 Males and 13 females). The infants were approximately 8.0 months old (range 7.5-8.5), with a mean age of 246 days (Range: 224-262; SD=9.4). The data from an additional 3 infants were excluded due to fussiness. Participants were recruited to come to the lab via letters mailed by the Maryland State Health Department, advertisements in local parenting magazines, and referrals from past participants.

---

[1] Although this explanation seems highly unlikely, an unpublished study carried out in the Jusczyk lab by Johnson and Jusczyk could possibly lend credence to this alternative explanation. Twenty-four English-learning 8-month-olds were familiarized with passages containing *tab* and *gondola* or *commuter* and *dresser*. All infants were tested on lists of the same four words: *tab*, *gondola*, *commuter*, and *dresser*. During the test phase, infants failed to listen significantly longer to familiarized test items. Johnson and Jusczyk attributed this null result to an overall preference to listen to the word *gondola* (Infants tended to listen longer to *gondola*, regardless of familiarization condition. The interaction between test item and group was in fact significant due to this behavior. In addition, adults naïve to the purpose of the study tended to pick out *gondola* as the most interesting test list.). However, it is also possible that the null result in this study might be due to the fact that words of different lengths were used during the test phase. If this were so, then the null result in the current study may have nothing to do with the complexity of the artificial language.

Parental consent was obtained for all participants. In appreciation for their participation, all infants were given a small toy or certificate and a travel reimbursement.

*Stimuli*

The same familiarization speech streams were used as in Experiment 1. The only difference between Experiment 1 and Experiment 2 is the use of test items of uniform rather than variable length. Infants familiarized with Language A were tested on the following four test items: *pabiku*, *tibudo*, *pigola*, and *latuda*. Infants familiarized with Language B were tested on a second set of four test items: *tudaro*, *pigola*, *pabiku*, and *kutibu*. Test items were concatenations of the individual syllables recorded for use in Johnson and Jusczyk (2001). Statistical part words were formed by taking the last syllable of one statistical word and combining it with two syllables from the beginning of another statistical word. All infants were tested on two statistical words and two statistical part words.

*Design*

All infants were randomly assigned to hear one of the two streams of speech (Language A or Language B). Twelve test trials were presented immediately after the familiarization phase. Half of the trials consisted of repetitions of trisyllabic statistical part words; the other half consisted of repetitions of trisyllabic statistical words.

*Procedure and Apparatus*

The procedure and apparatus were the same as Experiment 1.

<u>Results and Discussion</u>

Mean orientation times to the types of test items (statistical word versus statistical part word) during the test phase were calculated for each infant. Nine out of 24 infants had longer orientation times for the novel part-words. Across all infants, the average orientation times were 7.3 seconds (SD = 1.9) for the novel part-words, and 6.6 seconds (SD = 2.2) for the familiar words. A 2 X 2 mixed design ANOVA with Condition (familiarized with Language A versus Language B) and test item (statistical word versus statistical part-word) revealed no significant effect of test item, $F(1, 22) = 3.2$, $p > .05$. There was no significant effect of group, $F (1, 22) = < 1$, and no significant interaction between group and test item, $F(1, 22) = 1.3$, $p > .05$.

It appears that learning the language used in the current study was more difficult than learning the language used by Saffran et al. This suggests that the simplicity of the artificial language used by Saffran et al. played a major part in infants' success at tracking transitional probabilities. Even when the statistical structure of the language remains identical (1.0 between syllables within words and .33 between syllables spanning a word boundary), adding any amount of variation can make the language more difficult to learn. This result leads one to wonder if other types of variation, such as having a few more words in the language or using multiple different recordings of each syllable, would have a similar effect on the learnability of the artifiical language.

## General Discussion

The artificial language studies carried out be Saffran et al. (1996) and Aslin et al. (1998) are often cited as evidence that infants are extremely adept at learning the statistical structure of a patterned input in virtually no time at all. The two experiments carried out in this study suggest that infants' ability to learn the statistical structure of an artificial language can be severely hindered by merely varying the length of the words in the language. This simple manipulation did not change the strength of the transitional probablilies differentiating a within-word syllable sequence from a between-word syllable sequence. However, it did add an additional level of complexity to the artificial language used by Saffran et al. It seems unlikely that infants' failure in the current study is due to their inability to segment longer words, since there is ample evidence that 7.5-month-olds can segment bisyllabic and trisyllablic words from fluent speech (Houston, Santelman, and Jusczyk, submitted).

Infants' difficulty in learning the statistical structure of this slightly modified version of Saffran et al.'s language suggests that tracking the transitional probabilities between syllables in a real language (e.g., English) might be a bit more difficult than the results of Saffran et al.'s study first made it appear. Natural languages are highly variable, so if tracking the transitional probabilities between syllables is truly the sole key to getting infants' early segmentation capabilities off the ground, then we would expect infants to have no trouble segmenting segmenting a 4-word artificial language (even if the language contains words of variable length).

One could argue that that this is an unreal test for infants, given that they only had two minutes of exposure. But in fact, hearing the same four words repeating one after another for two minutes straight would never happen in a real language learning experience. Moreover, natural speech contains much more variation and ambiguity than the artificial speech stream used in the current study. Therefore, the process of word segmentation in the real world is undoubtedly much more complex than simply tracking the conditional probabilities between syllables. Although a computer may be able to find word boundaries in transcribed speech by tracking transitional probabilities between syllables, the results of Saffran et al's study give us no reason to believe that infants can segment fluent speech using a statistical strategy alone. First, tracking transitional probabilities between all syllables in the ambient language would be enormously taxing on the presumably small working memory of an infant. It seems more likely that infants track only highly salient or highly frequent patterns. Second, syllable boundaries are not always clearly marked in the speech signal. Even if syllable boundaries were easy to perceive, how would an infant decide what constitutes two tokens of the same syllable versus two tokens of two different syllables? Is a stressed token of the syllable *let* the same as an unstressed token (e.g. *let* in *bracelet* versus *lettuce*)? Is the highly aspirated *pa* in *pasta* the same as the *pa* in *grandpa*? Is the *you* in a rapidly produced token of *miss you* the same as the *you* in a carefully articulated occurrence of *miss you*? If tracking transitional probabilities are to be taken as a serious strategy for learning to segment the speech stream, then these sorts of factors must be taken into account. The input to infants is not a list of syllables. It is a complex acoustic pattern that requires a great deal of interpretation and re-encoding before syllable-sized units can be perceived. Given that syllables do not automatically emerge from the speech signal without a fair amount of processing, it seems clear that tracking transitional probabilities is not a full explanation for the onset of word segmentation.

The motivation underlying the current study is not to negate the finding that infants are highly adept at statistical learning. There is ample evidence that infants (and even primates) are quite good at statistical learning in multiple domains (Hauser, Newport, & Aslin, 2001; Saffran, Johnson, Aslin, & Newport, 1999). In addition, there is evidence that infants seem to pay attention to transitional probabilities when segmenting words from English passages (Jusczyk, Houston, & Newsome, 1999). However, the results of the current study suggest that infants' statistical learning capabilities might be of limited use in a natural language setting. Thus, we must be cautious before concluding that statistical learning explains the bulk of how infants begin segmenting words. Rather, we must begin to ask how infants' statistical learning prowess could be realistically applied in the domain of early word segmentation.

References

Aslin, R.N., Saffran, J.R., & Newport, E.L. (1998). Computation of conditional probability statistics by 8-month-old infants. Psychological Science, 72, 321-324.

Bruner, J.S. (1973). The identification of recurrent regularity, (pp. 198-207). In: Beyond the Information Given. NY: Norton and Company.

Hauser, M., Newport, E.L., Aslin, R.N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in Cottop-Top Tamarins. Cognition,

Houston, D.M., Santalman, L., Jucczyk, P.W. (submitted). English-learning infants segmentation of trisyllabic words from fluent speech. Language and Cognitive Processes.

Hunter, M.A. & Ames, E.W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. IN C. Rovee-Collier & L.P. Lipsitt (Eds.). Advances in Infancy Research (Volume 5, pp. 64-95). Norwood, NJ: Ablex.

Johnson, E.K. & Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. Journal of Memory and Languag, 44, 548-567.

Jusczyk, P.W. (1997). The discovery of Spoken Language. Cambridge, MA: MIT Press.

Jusczyk, P.W., Houston, D.M., and Santelman, L. (1999). The beginnings of word segmentation in English-learning infants. Cognitive Psychology, 39, 159-207.

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.

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. (1999). Statistical learning of tone sequences by human adults and infants. Cognition, 70, 27-52.