Research Report

COMPUTATION OF CONDITIONAL PROBABILITY STATISTICS BY 8-MONTH-OLD INFANTS

Richard N. Aslin, Jenny R. Saffran, and Elissa L. Newport
Department of Brain and Cognitive Sciences, University of Rochester

Abstract—*A recent report demonstrated that 8-month-olds can segment a continuous stream of speech syllables, containing no acoustic or prosodic cues to word boundaries, into wordlike units after only 2 min of listening experience (Saffran, Aslin, & Newport, 1996). Thus, a powerful learning mechanism capable of extracting statistical information from fluent speech is available early in development. The present study extends these results by documenting the particular type of statistical computation—transitional (conditional) probability—used by infants to solve this word-segmentation task. An artificial language corpus, consisting of a continuous stream of trisyllabic nonsense words, was presented to 8-month-olds for 3 min. A postfamiliarization test compared the infants' responses to words versus part-words (trisyllabic sequences spanning word boundaries). The corpus was constructed so that test words and part-words were matched in frequency, but differed in their transitional probabilities. Infants showed reliable discrimination of words from part-words, thereby demonstrating rapid segmentation of continuous speech into words on the basis of transitional probabilities of syllable pairs.*

Many aspects of the patterns of human languages are signaled in the speech stream by what is called distributional evidence, that is, regularities in the relative positions and order of elements over a corpus of utterances (Bloomfield, 1933; Maratsos & Chalkley, 1980). This type of evidence, along with linguistic theories about the characteristics of human languages, is what comparative linguists use to discover the structure of exotic languages (Harris, 1951). Similarly, this type of evidence, along with tendencies to perform certain kinds of analyses on language input (Chomsky, 1957), could be used by human language learners to acquire their native languages. However, using such evidence would require rather complex distributional and statistical computations, and surprisingly little is known about the abilities of human infants and young children to perform these computations. By using the term computation, we do not mean, of course, that infants are consciously performing a mathematical calculation, but rather that they might be sensitive to and able to store quantitative aspects of distributional information about a language corpus.

Recently, we have begun studying this problem by investigating the abilities of human learners to use statistical information to discover word boundaries (Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; for prior work on this topic, see also Goodsitt, Morgan, & Kuhl, 1993; Harris, 1955; Hayes & Clark, 1970). Words are known to vary dramatically from one language to another, so finding the words of a language is clearly a problem that must involve learning from the linguistic environment. Moreover, the beginnings and ends of the sequences of sounds that form words in a particular language are not marked by any consistent acoustic cues (Aslin, Woodward, LaMendola, & Bever, 1996; Cole & Jakimik, 1980; Lehiste, 1970), so the problem must be solved, at least in part, by some type of distributional, rather than acoustic, analysis.

We recently provided the first evidence that 8-month-old infants can use sequential statistics to group sounds into wordlike units (Saffran, Aslin, & Newport, 1996). In that study, we created two small artificial languages consisting of "words" that were each composed of three nonsense syllables. We then examined the ability of human infants to learn to segment the languages into these wordlike units after very brief exposure to corpora that contained uninterrupted sequences of the words. Two different artificial languages were created by combining 12 different syllables to form four trisyllabic nonsense words for each language. The two languages differed only in which particular order of the syllables was used to form the words; using these two languages therefore controlled for the possibility that particular sequences of syllables might form units more readily than others. For each language, the words were presented in random order, with the constraint that no word was repeated in immediate succession, to form a continuous 2-min speech stream. A speech synthesizer was programmed to speak this continuous stream with no pauses or pitch or duration changes at the word boundaries. Although there were no acoustic cues to word boundaries, there was nonetheless statistical information that could allow the identification of the words from which the corpus was formed. Our experimental question was whether infants would be able to utilize such information to discriminate those syllable sequences that formed words from those syllable sequences that occurred in the corpus but did not form words.

In each artificial language, each syllable occurred in only one of the four words, and in a unique position in that word. Thus, the *transitional probability* between successive syllables X and Y, a type of conditional probability statistic (Miller & Selfridge, 1950), defined as

$$\text{probability of } Y|X = (\text{frequency of } XY) \,/\, (\text{frequency of } X),$$

was 1.0 for syllable pairs internal to words. In contrast, the transitional probabilities between successive syllables that happened to occur next to each other across word boundaries were lower. In real languages, of course, a particular syllable appears in many different words. Word-internal transitional probabilities are thus below 1.0, but still higher than transitional probabilities across word boundaries.[1]

It is important to note that in acquisition of natural rather than artificial languages, distributional analysis of language input is unlikely to

Jenny R. Saffran is now at the Department of Psychology, University of Wisconsin-Madison.

Address correspondence to Richard N. Aslin, Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, NY 14627; e-mail: aslin@cvs.rochester.edu.

1. There are other conditional probability statistics (e.g., conditional entropy, mutual information, correlation) that are functionally equivalent to transitional probabilities, in that they all normalize co-occurrence frequency by the overall frequency of individual events. Any of these conditional probability statistics, including backward transitional probability (the probability of X given Y), provides information for word segmentation on the basis of low predictability at word boundaries.

rely on a single type of information (see Brent & Cartwright, 1996; Christiansen, Allen, & Seidenberg, in press), such as frequency of co-occurrence or transitional probability between adjacent syllables. Rather, natural word segmentation likely relies on a variety of sources of information, including prosody (Cutler & Norris, 1988; Jusczyk, Cutler, & Redanz, 1993); pauses and prosodic changes at syllable, phrase, and utterance boundaries (Gleitman & Wanner, 1982; Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Morgan, Meier, & Newport, 1987; Morgan & Newport, 1981; Morgan & Saffran, 1995; Myers et al., 1996); and phonotactics (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993), even though none of these sources of information alone is reliable enough in natural speech input to solve the word-segmentation problem. Moreover, distributional analyses can contribute to the organization of language at levels higher than the word only if they involve rather different statistical computations than the ones examined here. Thus, our studies of word segmentation are intended to provide an indication of the power of distributional analyses for sequence learning, rather than to demonstrate that these statistical learning mechanisms account for all of language acquisition.

Our focus in the present study was to identify the particular type of statistic infants utilize to segment words from fluent speech. We were especially interested in the degree of computational complexity they bring to this, and related, sequence-learning tasks. The critical manipulation in the present study involved equating the frequencies of the syllable sequences within versus across words, while maintaining the differences in transitional probabilities. If each of the four words in the language appeared equally often in the speech stream, the trisyllabic sequences that form the words would not only have higher transitional probabilities than trisyllabic sequences across word boundaries (part-words), they would also occur more frequently. This was true of our previous study. Our results in that study definitively showed that infants could discriminate words from part-words, thus demonstrating that they were sensitive to some rather impressive statistical property (either transitional probabilities among three-syllable sequences or the frequency of co-occurrence of the syllables in these sequences). However, given its design, that study could not discern which of these two types of statistical analyses was actually performed.

In the present study, two of the words occurred twice as often in the corpus as the other two. As a result, the syllable sequences across boundaries between the two common words occurred with relatively high frequency. In fact, the frequency of the trisyllabic part-words formed in this fashion was precisely equal to the frequency of the less common trisyllabic words. But equating the frequency of the words did not equate the transitional probabilities of their component syllables. The two syllable pairs within a word still had transitional probabilities of 1.0, whereas the transitional probabilities of the two syllable pairs within these high-frequency part-words were .5 (across a word boundary) and 1.0.[2] Thus, the design of the present study enabled us to determine if infants could rely solely on transitional probabilities to segment words from fluent speech. We predicted that evidence of such segmentation would result in longer listening times to the part-words than to the words, a novelty effect observed in both of the experiments reported in Saffran, Aslin, and Newport (1996).

2. This test was also somewhat more difficult than the test in our previous study in that the transitional probabilities across word boundaries were higher (.50, compared with .33).

METHOD

Subjects

Two groups of fifteen 8-month-olds were tested (mean age = 8 months, 0 weeks; range: 7 months, 2 weeks to 8 months, 2 weeks). An additional 7 infants did not complete the familiarization and testing phases because of fussiness (6) or drowsiness (1). All infants were solicited from local birth announcements and hospital records, and parental consent was obtained prior to testing in accordance with the guidelines of the local human subjects review committee and the principles of ethical treatment as established by the American Psychological Association.

Stimuli

All stimuli were generated by the MacinTalk© speech synthesizer, edited to equate syllable durations, and stored on disk at a sampling rate of 22 kHz for on-line playback through an Audiomedia soundboard in a Quadra 650 computer. The four trisyllabic nonsense words in Corpus 1 were *pabiku, tibudo, golatu,* and *daropi,* and the four words in Corpus 2 were *tudaro, pigola, bikuti,* and *budopa.* Each 3-min corpus consisted of 270 trisyllabic word tokens, produced by the synthesizer at a rate of 4.5 syllables/s, with no pauses or other acoustic cues to word boundaries. (An example from one corpus, with orthographic continuity used to indicate the lack of acoustic cues to word boundaries, is as follows: *pabikugolatudaropitibudodaropigolatu. . . .*) The first two words in the list for each corpus occurred 45 times each, and the second two words occurred 90 times each.

The four test items were *pabiku, tibudo, tudaro,* and *pigola.* The first two test items were words in Corpus 1 and part-words in Corpus 2. The second two test items were part-words in Corpus 1 and words in Corpus 2. Note that part-words consisted of the final syllable of one word and the first two syllables of another word. These words that formed the part-words occurred more frequently in the familiarization corpus. All four test items, the two words and the two part-words, occurred 45 times each in both familiarization corpora.

Procedure

Each infant was tested individually while seated in a parent's lap in a sound-attenuated booth. An observer outside the booth monitored the infant's looking behavior on a closed-circuit television system and coded the infant's behavior using a button-box connected to the computer. This button-box was used to initiate trials and to enter the direction of the infant's head turns, which controlled the duration of each test trial. Both the parent and the observer listened to masking music over headphones to eliminate bias.

At the beginning of the 3-min familiarization phase, the infant's gaze was first directed to a blinking light on the front wall in the testing booth. Then the sound sequence for one of the two corpora was presented without interruption from two loudspeakers, one located on each of the two side walls of the booth. During this familiarization period, in an effort to keep the infant's interest, a blinking light above one of the two loudspeakers (randomly selected) was lit and extinguished dependent on the infant's looking behavior, but there was no contingency between lights and sound, which played continuously.

Immediately after familiarization, 12 test trials were presented (3 trials for each of the four test items, presented in random order). Six of these trials were thus words, and 6 were part-words. Each test trial began with the blinking light on the front wall. When the observer signaled the computer that the infant was fixating this central light, one of the lights on the two side walls began to blink, and the central light was extinguished. When the observer judged that the infant had made a head turn of at least 30° in the direction of the blinking side light, a button-press signaled to the computer that one of the trisyllabic test items should be presented from the loudspeaker adjacent to the blinking light. This test item was repeated, with a 500-ms silent interstimulus interval, until the observer coded the infant's head turn as deviating away from the blinking light for 2 consecutive seconds. When this look-away criterion was met, the computer extinguished the blinking side light, turned off the test stimulus, and turned on the central blinking light to begin another test trial. The computer randomized the order of test trials and accumulated total looking time to each of the two test words and two part-words.

RESULTS

Looking times for words and for part-words were averaged across the two groups because no differences were observed between the two artificial languages, $t(28) = 0.63$, $p > .05$. As in our previous study (Saffran, Aslin, & Newport, 1996), infants showed a significant difference in listening times to the two types of test items, $t(29) = 2.10$, $p < .05$, with longer listening times to the part-words than to the words (see Table 1). This difference shows that after only 3 min of listening to the language corpus, the infants were able to discriminate sequences of syllables that formed words from sequences of syllables that were closely matched in many regards, but differed in their transitional probabilities.

Our interpretation of this finding is that infants can discriminate the differences in transitional probabilities between words and part-words, and that (as in our two previous experiments using nearly identical stimuli and testing procedures) they prefer to listen to the relatively novel part-words over the relatively familiar words. Recall that the transitional probabilities of the two syllable pairs within words were 1.0 and 1.0, whereas the transitional probabilities for part-words were .5 and 1.0. There is, however, an alternative interpretation based not on transitional probabilities, but rather on syllable frequencies. Although our words and part-words were equated for the frequency of co-occurrence of their three syllables, they differed in the frequency of the syllables and bisyllables of which these trisyllables were composed. The two syllable pairs in the test words occurred 45 times each in the familiarization corpus, whereas the two syllable pairs in the part-words occurred 45 and 90 times, respectively. However, if infants' longer listening times to the part-words were based on the higher frequency of occurrence of a syllable pair in the part-words, then these listening times would be indicative of a familiarity (rather than a novelty) effect.[3] Such an effect would be quite surprising because, as already noted, our earlier studies using this paradigm consistently showed a novelty effect (see Table 1). That is, in our earlier studies, the longer listening times to nonwords or part-words could not have been a familiarity effect because all relevant aspects of these stimuli—their bisyllable frequencies and their transitional probabilities—were less familiar than those of the words. Thus, unless one assumes that listening preferences vary randomly across experiments of the same type, the pattern of results across these studies argues rather strongly that our original interpretation, and not this alternative, is correct: Infants in the present study discriminated between words and part-words on the basis of differences in transitional probabilities.

Table 1. *Mean listening times (in seconds) to the test items in the present study and two previous experiments*

Study	Words	Part-words/ Nonwords
Present study	6.78 (0.36)	7.36 (0.42)
Saffran, Aslin, and Newport (1996), Experiment 1	7.97 (0.41)	8.85 (0.45)
Saffran, Aslin, and Newport (1996), Experiment 2	6.77 (0.44)	7.60 (0.42)

Note. Standard errors are shown in parentheses.

DISCUSSION

These results extend and greatly sharpen our previous report of a rapid statistical learning mechanism in young infants. By design, the words and part-words in the present study were equated in the frequency with which their syllables occurred together, in that order, in the familiarization corpus. Only the difference in transitional probabilities between successive syllables, and not the frequency of syllable co-occurrence, can account for these results. Evidence of discrimination between words and part-words when trisyllabic frequencies were equated indicates that 8-month-olds are sensitive to the conditional probabilities of successive sounds (syllables) and can organize sound sequences on the basis of differences in these conditional probabilities.

The computation of conditional probabilities is an important ability because, in language as in many other patterned domains, relative frequency (even complex frequency, such as the frequency of co-occurrence of pairs or triples of items) is not the best indicator of structure. Instead, significant structure is typically most sharply revealed by the statistical predictiveness among items (i.e., frequency of co-occurrence normalized for frequency of the individual components; see Rescorla, 1966).[4] For

3. Although familiarity effects have been obtained in several studies of language preferences (e.g., Jusczyk & Aslin, 1995), studies using stimuli and familiarization procedures closest to our own have reported novelty effects (e.g., Echols, Crowhurst, & Childers, 1997).

4. Rescorla (1966) showed that classical conditioning in dogs involves the computation of a conditional probability or correlation between a tone and subsequent presentation of shock. One might ask, then, if human infants can show classical conditioning, is it not already known that they can compute conditional probabilities? In fact, to our knowledge, Rescorla's paradigm has not been run with human infants. But, more important, our own task involves quite a different order of magnitude of processing than Rescorla's. Our word-segmentation task, if performed in its entirety, involves the on-line (running) computation of 20 different conditional probabilities, each over 45 to 90 occurrences of the component syllables and 9 to 90 occurrences of syllable pairs, during a 3-min learning period. Eight of these 20 conditional probabilities are included in our test items. Our study thus asks not merely whether infants can compute a single conditional probability, but whether they can compute a large number of such probabilities simultaneously.

example, in English, it is common for the determiner *the* to appear immediately prior to nouns. If a particular noun were highly frequent in a corpus, an algorithm that segmented words based on co-occurrence frequency would mistakenly undersegment *the* + *noun*. However, the presence of several different nouns following the determiner *the* would, by a transitional probability algorithm, appropriately segment *the* + *noun*. It remains to be seen whether infants can perform similar analyses of nonlinguistic materials,[5] or whether this computational ability is restricted to the language domain, in which humans have evolved a number of special adaptations (e.g., Lieberman, 1984; Pinker & Bloom, 1990).

Our findings also raise questions about the range of statistical analyses infants are able to perform on linguistic materials, and about the ways in which these computations are integrated with other types of linguistic information in the process of learning a language. It has been known for many years that word segmentation could in principle be accomplished, at least in part, by computing transitional probabilities across successive sounds (Harris, 1955; Hayes & Clark, 1970). What was not known was whether young language learners were capable of performing such computations. Other aspects of language (e.g., syntax), however, cannot be captured only by statistical analyses that operate on adjacent items (Chomsky, 1957). Explaining the acquisition of these higher levels of language structure therefore requires either a quite different type of mechanism or even more complex distributional analyses and richer representations over which the analyses are performed (Mintz, Newport, & Bever, 1995; Saffran, 1997). Although we do not yet have answers to these important questions, we believe the present results suggest that human infants possess a larger stock of complex computational abilities than previously believed.

Acknowledgments—We thank J. Gallipeau and J. Hooker for their help with various aspects of this research. Support was provided by a National Science Foundation research grant to R.N.A. (SBR-9421064), a National Science Foundation predoctoral fellowship to J.R.S., and a National Institutes of Health research grant to E.L.N. (DC00167).

REFERENCES

Aslin, R.N., Woodward, J., LaMendola, N., & Bever, T.G. (1996). Models of word segmentation in fluent maternal speech to infants. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). Mahwah, NJ: Erlbaum.

Bloomfield, L. (1933). *Language.* New York: Holt, Rinehart & Winston.

Brent, M.R., & Cartwright, T.A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93–120.

Chomsky, N. (1957). *Syntactic structures.* The Hague, the Netherlands: Mouton.

Christiansen, M.H., Allen, J., & Seidenberg, M.S. (in press). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes.*

Cole, R.A., & Jakimik, J. (1980). A model of speech perception. In R.A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Erlbaum.

5. There is certainly prior evidence of learning and memory in young infants (see, e.g., Rovee-Collier & Sullivan, 1980) and of categorization based on correlations among attributes (see, e.g., Younger, 1985). However, we know of no previous studies of infant learning in which elements presented in rapid succession (4/s) were organized on the basis of a complex, multielement, conditional probability analysis conducted over a very brief (3 min) familiarization period. Preliminary evidence (Johnson, Saffran, Aslin, & Newport, 1998) suggests that nonlinguistic (tonal) stimuli can also be grouped on the basis of transitional probabilities.

Cutler, A., & Norris, D.G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113–121.

Echols, C.H., Crowhurst, M.J., & Childers, J.B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language, 36,* 202–225.

Friederici, A.D., & Wessels, J.M.I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics, 54,* 287–295.

Gleitman, L.R., & Wanner, E. (1982). Language acquisition: The state of the art. In E. Wanner & L.R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, England: Cambridge University Press.

Goodsitt, J.V., Morgan, J.L., & Kuhl, P.K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language, 20,* 229–252.

Harris, Z.S. (1951). *Methods in structural linguistics.* Chicago: University of Chicago Press.

Harris, Z.S. (1955). From phoneme to morpheme. *Language, 31,* 190–222.

Hayes, J.R., & Clark, H.H. (1970). Experiments in the segmentation of an artificial speech analog. In J.R. Hayes (Ed.), *Cognition and the development of language* (pp. 221–234). New York: Wiley.

Hirsh-Pasek, K., Kemler Nelson, D.G., Jusczyk, P.W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269–286.

Johnson, E.K., Saffran, J.R., Aslin, R.N., & Newport, E.L. (1998, April). *Statistical learning of non linguistic stimuli by 8-month-olds.* Poster presented at the International Conference on Infancy Studies, Atlanta.

Jusczyk, P.W., & Aslin, R.N. (1995). Infants' detection of the 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., Friederici, A.D., Wessels, J., Svenkerud, V.Y., & Jusczyk, A.M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32,* 402–420.

Jusczyk, P.W., Hirsh-Pasek, K., Kemler Nelson, D.G., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24,* 252–293.

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

Lieberman, P. (1984). *The biology and evolution of language.* Cambridge, MA: Harvard University Press.

Maratsos, M., & Chalkley, M.A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), *Children's language* (Vol. 2, pp. 127–213). New York: Gardner Press.

Miller, G.A., & Selfridge, J.A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63,* 176–185.

Mintz, T.H., Newport, E.L., & Bever, T.G. (1995). Distributional regularities of form class in speech to young children. In J.N. Beckman (Ed.), *Proceedings of NELS 25* (Vol. 2, pp. 43–54). Amherst, MA: Graduate Linguistic Student Association.

Morgan, J.L., Meier, R.P., & Newport, E.L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19,* 498–550.

Morgan, J.L., & Newport, E.L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior, 20,* 67–85.

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

Myers, J., Jusczyk, P.W., Kemler Nelson, D.G., Charles-Luce, J., Woodward, A., & Hirsh-Pasek, K. (1996). Infants' sensitivity to word boundaries in fluent speech. *Journal of Child Language, 23,* 1–30.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioural and Brain Sciences, 13,* 707–784.

Rescorla, R.A. (1966). Predictability and number of pairings in Pavlovian fear conditioning. *Psychonomic Science, 4,* 383–384.

Rovee-Collier, C., & Sullivan, M.W. (1980). Organization of infant memory. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 798–807.

Saffran, J.R. (1997). *Statistical learning of syntactic structure: Mechanisms and constraints.* Unpublished doctoral dissertation, University of Rochester, Rochester, NY.

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

Saffran, J.R., Newport, E.L., & Aslin, R.N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606–621.

Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8,* 101–105.

Younger, B. (1985). The segregation of items into categories by ten-month-old infants. *Child Development, 56,* 1574–1583.

(RECEIVED 8/25/97; ACCEPTED 12/10/97)