INFANCY, *1*(1), 123–147

RESEARCH ARTICLE

Six-Month-Olds' Detection of Clauses Embedded in Continuous Speech: Effects of Prosodic Well-Formedness

Thierry Nazzi
Institute of Human Development
University of California at Berkeley

Deborah G. Kemler Nelson
Department of Psychology
Swarthmore College

Peter W. Jusczyk
Department of Psychology and Cognitive Science
Johns Hopkins University

Ann Marie Jusczyk
Department of Psychology
Johns Hopkins University

Three experiments investigated the role of prosodic structure for infants' recognition of embedded word sequences. Six-month-olds were familiarized with 2 versions of the same sequence, 1 corresponding to a well-formed prosodic unit and the other to a prosodically ill-formed sequence (although a successive word series). Next, infants

Supplementary materials to this article are available on the World Wide Web at http://www.infancyarchives.com.

Requests for reprints should be sent to Thierry Nazzi, Institute of Human Development, University of California at Berkeley, 1203 Tolman Hall, MC 1690, Berkeley, CA 94720–1690, or Peter W. Jusczyk, Department of Psychology, Ames Hall, Johns Hopkins University, Baltimore, MD 21218–2686.

heard 2 test passages. One included the well-formed unit, and the other included the ill-formed sequence. In Experiment 1, infants listened longer to the passage containing the well-formed unit, suggesting that such units, even when they are embedded, are better recognized. Experiments 2 and 3 showed that this better recognition does not depend on an acoustic match between the familiarized sequences and their later embeddings. This suggests that the advantage of the well-formed unit is at least partially due to infants' use of prosody to parse continuous speech.

An important part of learning any language is to discover the syntactic organization that underlies utterances. A preliminary step in this process is to segment the input in such a way that it preserves the important units determined by the grammar, such as clauses. How does a learner carve up the input in a way that allows for the recovery of the units of syntactic organization? Advocates of a prosodic bootstrapping account have suggested that learners may use cues available in the speech signal as a starting point for segmenting the input into the appropriate syntactic units (Gleitman & Wanner, 1982; Morgan, 1986; Peters, 1983). Indeed, there is considerable evidence indicating that the speech signal does contain reliable cues to the boundaries of clausal units (Beckman & Edwards, 1990; Cooper & Paccia-Cooper, 1980; Klatt, 1975; Lehiste, Olive, & Streeter, 1976; Nakatani & Dukes, 1977; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Scott & Cutler, 1984). Moreover, such cues seem to be, if anything, exaggerated in speech directed to young language learners (Bernstein Ratner, 1986; Fisher & Tokura, 1996a).

Although the situation with respect to the marking of subclausal units such as major phrases is more complicated, there are indications that some acoustic marking of syntactic phrases does occur in speech to adults (Beach, 1991; Beckman & Edwards, 1990; Price et al., 1991; Scott & Cutler, 1984) and to infants (Jusczyk, Hirsh-Pasek, et al., 1992; Lederer & Kelly, 1991). True, the marking of such units is not always present or consistent (Ferreira, 1993; Fisher & Tokura, 1996b; Gee & Grosjean, 1983; Gerken, Jusczyk, & Mandel, 1994). The phrasal units that are typically picked out in the input are usually prosodic phrases, and these do not map consistently to particular syntactic phrases (Nespor & Vogel, 1986; Selkirk, 1984). Still, they usually do map to a syntactic phrase of some sort (Fisher, in press; Gerken et al., 1994; Jusczyk, 1998a, 1999). Overall then, there is some information available in the speech signal that could help the learner break up the input into the kinds of units that are important for learning about the syntactic organization of language.

Given that some marking of syntactic units in speech exists, what evidence is there that infants are sensitive to such information? Results of a number of investigations indicate that infants listening to infant-directed speech are attuned to acoustic cues that differentiate boundaries of major syntactic units, such as clauses and phrases, from nonboundary locations. Much of this evidence comes from studies using a pause-insertion manipulation. Two versions of otherwise natural infant-directed speech are created by inserting artificial pauses at unit boundaries or at within-unit locations. Infants under 1 year of age have been shown to listen longer to samples interrupted at clause boundaries or major phrase boundaries than to samples interrupted in the middle of such units (Hirsh-Pasek et al., 1987; Jusczyk, Hirsh-Pasek, et al., 1992; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989). Indeed, for English clausal units, the preference can be detected as early as 4½ months of age (Jusczyk, 1989; Kemler Nelson, 1989), and for English phrasal units, it appears by 9 months (Jusczyk, Hirsh-Pasek, et al., 1992). Importantly, low-pass filtering of the speech samples, which removes most phonetic information while preserving most prosodic information, leaves the preference intact (Jusczyk, 1989; Jusczyk, Hirsh-Pasek, et al., 1992). Accordingly, in distinguishing pauses at boundary and nonboundary locations, infants appear to be tuning in to prosodic information that is potentially useful for segmenting fluent speech in syntactically relevant ways.

Such results are not peculiar to the pause-insertion technique. Morgan, Swingley, and Miritai (1993) reported parallel preferences for speech segmented at clause boundaries when buzzing noises rather than artificial pauses were inserted. Using a different methodology altogether, Morgan (1994) also showed that infants were better at detecting a noise when it occurred at clause boundaries than when it occurred between words of the same clause. Thus, infants not only prefer an artificial segmentation (by pauses or noises) that coincides with clause boundaries, but they are more responsive to extraneous interruptions when they occur at these boundaries. These findings provide converging evidence of infants' attention to prosodic cues that are potentially useful for processing syntactic units and learning language.

Is there any evidence that young infants' sensitivity to prosodic cues actually plays a role in how they process and organize natural speech? Does their ability to detect potentially relevant prosodic cues have real consequences for how infants represent the fluent speech that they hear? Such evidence is required if the prosodic bootstrapping account is to be given very serious consideration.

A first step toward addressing the issue was taken by Mandel and colleagues (Mandel, Jusczyk, & Kemler Nelson, 1994; Mandel, Kemler Nelson, & Jusczyk, 1996). They reasoned that if prosodic cues help infants organize their representations of incoming speech, it should be possible to show that prosodic structure influences how infants encode and remember speech. In particular, the packaging of speech into a well-formed prosodic unit might benefit even very young infants' memory for the linguistic information it contains. Using the high-amplitude sucking paradigm, Mandel et al. (1994) tested 2-month-old infants' memory for the phonetic properties of words linked prosodically in a single clause versus their memory for the same information either when it was spoken as (a) a list of disconnected words or (b) two contiguous and prosodically incomplete fragments. In both cases, the well-formed prosodic organization gave the infants a phonetic memory advantage. Mandel et al. (1996) demonstrated further that the coherence provided by a single, well-formed clause also affords infants better memory for word-order information. Two-month-olds retained the order of words better (i.e., they dishabituated to a change in word order) when the words were spoken as a single, well-formed clause rather than as two different clausal fragments. Together, these are the first indications that infants are not only sensitive to prosodic cues to linguistic organization but that they use prosodic information to help them represent the speech they hear.

The effects observed by Mandel and colleagues (Mandel et al., 1994; Mandel et al., 1996), impressive as they are as evidence of infants' use of prosody in encoding speech, are somewhat limited in scope. In particular, their findings indicate that a single, well-formed prosodic envelope, relative to ill-formed prosodic structure, enhances the memorability of the speech that it contains. However, the well-formed prosodic speech used during both the habituation and test phases of Mandel and colleagues' studies consisted of only a single, short spoken clause rather than a connected passage of prosodically well-formed clauses. What remains to be seen is whether the natural prosody of a clause also allows infants to bracket or segment continuous speech in a way that permits the recognition of a prosodically well-formed clause embedded in a longer passage. Arguments for the usefulness of prosodic information in early language acquisition would be significantly strengthened if infants were shown to use prosody to recognize linguistically relevant units that are embedded in continuous speech. Our investigations constitute the beginning of an attempt to seek such evidence.

In three experiments, 6-month-old infants were familiarized with sequences of words in isolation and then tested for their ability to recognize those sequences when they were embedded in longer passages. During familiarization and test, the same word sequences were spoken both as a single, prosodically well-formed clause and as a concatenated series of prosodically ill-formed segments, that is, word sequences that began and ended in the middle of clauses and that crossed a clause boundary. If the prosodic envelope of a spoken clause helps infants to detect it when it is embedded in continuous speech, infants should be more likely to recognize a prosodically well-formed reoccurrence of the clause than a reoccurrence of a prosodically ill-formed sequence of the same words.

EXPERIMENT 1

A notable limitation of the high-amplitude sucking procedure is that it only permits the presentation of relatively brief stimuli, usually no longer than 1.5 sec in duration (Polka, Jusczyk, & Rvachew, 1995). Thus, although this procedure can be used to present a single spoken clause, as Mandel and colleagues (Mandel et al., 1994; Mandel et al., 1996) did, it is not suitable for presenting the kinds of multiple-clause passages necessary for exploring infants' abilities to recognize material embedded in extended speech. By comparison, the headturn preference procedure (HPP) is a method that allows for the presentation of relatively long excerpts of continuous speech (Jusczyk, 1998b; Kemler Nelson et al., 1995). Moreover, HPP has been a useful method for exploring word segmentation abilities of infants. For example, infants familiarized with a pair of isolated target words listened significantly longer to passages containing these targets than to ones without them (Jusczyk & Aslin, 1995; Jusczyk, Hohne, & Bauman, 1999; Jusczyk, Houston, & Newsome, 1999). Might infants who are familiarized with a prosodically well-formed clause subsequently listen significantly longer to a passage containing this clause than to one that does not include the clause as a well-formed unit? Is familiarization with a prosodically well-formed clause better than familiarization with a prosodically ill-formed sequence of the same words when it comes to recognizing the same sequence embedded in a passage? In this experiment, we addressed these questions by familiarizing 6-month-olds with both a prosodically well-formed sequence (corresponding to a clause) and a prosodically ill-formed sequence (from fragments of two clauses) and then testing whether they listened longer to a passage with the well-formed sequence than to one with the ill-formed sequence.

Method

Participants. Thirty-six American infants (21 girls, 15 boys) from monolingual English-speaking families were tested. The infants had an average age of 26 weeks, 6 days (range = 24 weeks, 0 days to 29 weeks, 2 days), or approximately 6 months. Eight additional infants were tested but not included for the following reasons: becoming fussy and crying (4), average orientation times shorter than 3 sec (3), and not turning to the lights (1).

Stimuli. The stimuli consisted of two passages and four short sequences. A female native speaker of American English recorded the following two passages:

(I) John doesn't know what *rabbits eat.* ***Leafy vegetables*** **taste so good.** They don't cost much either.
(II) Many animals prefer some things. **Rabbits eat *leafy vegetables.*** *Taste so good* is rarely encountered.

The four short sequences were extracted from these passages. Two of them were well-formed clauses (appearing in bold); the other two, the prosodically ill-formed sequences, came from adjacent portions of two different clauses (appearing in italics). The short sequences were presented during the familiarization phase and the passages during the test phase. The average duration of each short sequence was 2.1 sec. The average duration of each three-sentence passage was 7.66 sec. One sound file was constructed for each of the short sequences and for each of the passages. Each file was made up of several repetitions of the short sequences or passages (seven and three repetitions, respectively). Thus, the maximum trial duration of the familiarization sequences was 22.6 sec, whereas the maximum trial duration of the test passages was 25 sec. The sound files for these stimuli and all the stimuli used in other experiments in this article are available on the World Wide Web at http://www.infancyarchives.com.

Acoustic analyses conducted on the short sequences indicated that the well-formed and ill-formed sequences differed in the expected ways as predicted by the prosodic structure for clauses. For example, the ill-formed sequences had a large rise in pitch and an increase in root mean square (rms) amplitude between the second and third words of the sequences (e.g., *eat* and *leafy* for Passage I and *vegetables* and *taste* for Passage II). Silent gaps of 456 and 140 msec occurred between these two words in Passages I and II, respectively (the comparable values for the well-formed sequences were 27 and 36 msec). In addition, there was evidence that the final stressed syllable prior to the sentential break had an increased duration in each of the ill-formed samples. By comparison, the well-formed sequences showed no such changes in any of these measures in the same vicinity. Instead, the pitch peaks tended to fall on the initial word of each sequence and a secondary peak evident on the initial syllable of the final word of each sentence (*good* in Passage I and *vegetables* in Passage II). In each of these sentences, the final word received emphatic stress, and thus its final stressed syllable had both a longer duration and increased rms amplitude relative to the preceding word.

Design. The short sequences that were heard during the familiarization phase were counterbalanced across infants. For this purpose, infants were randomly assigned to one of two groups, defined in terms of the set of sequences (Set 1, the well-formed and ill-formed versions of “Leafy vegetables taste so good,” or Set 2, the two versions of “Rabbits eat leafy vegetables”) that they heard during the familiarization phase. During the test phase, both groups heard the same two test passages. However, which of the two test passages matched the well-formed and ill-formed familiarization sequences was reversed for the two groups. There were 18 infants in each group.

Apparatus. A PDP 11/73 computer controlled the presentation of the stimuli and recorded the observer’s coding of the infants’ responses. The stimuli were stored in digitized form (at a 20-kHz sampling rate) on the computer. A 12-bit Digital to Analog converter (at a 20-kHz sampling rate with a low-pass filter cutoff at 9.5 kHz) was used to re-create the audio signal. The output was fed through anti-aliasing filters and a Kenwood audio amplifier (KA 5700) to 7-in. Advent loudspeakers mounted behind the side walls of the testing booth. The stimuli in this experiment (and all subsequent experiments) were played at an amplitude of 72 ± 3 db (C) sound pressure level, about 15 db above the ambient noise level in the test room.

The experiment was conducted in a three-sided test booth made of pegboard panels. Except for a small section of preexisting holes in the front panel used for monitoring the infant's headturns, the panels were backed with white cardboard to prevent the infant from seeing behind them. The test booth had a red light and a loudspeaker (7-in. Advent) mounted on each of the side panels and a green light mounted on the center panel, all approximately at the seated infant's eye level. Directly below the center light, a 5-cm hole accommodated the lens of a video camera used to record each test session. A white curtain suspended around the top of the booth shielded the infant's view of the rest of the room. A computer terminal and response box were located behind the center panel, out of view of the infant. The response box, which was connected to the computer, was equipped with a series of buttons that started and stopped the flashing center and side lights, recorded the direction and duration of headturns, and terminated a trial when the infant looked away for more than 2 sec. Information about the direction and duration of the headturns and the total trial duration was stored in a data file on the computer.

Procedure. A modified version of the HPP was used in this study. Each infant was held on a caregiver's lap. The caregiver was seated on a chair in the center of the test booth. Each trial began by flashing the green light on the center panel until the infant had oriented in that direction. Then, the center light was extinguished and the red light above the loudspeaker on one of the side panels began to flash. When the infant made a turn of at least 30° in the direction of the loudspeaker, the stimulus for that trial began to play. It was played until its completion (i.e., when the entire sound file had been presented) or until the infant failed to maintain the 30° headturn for 2 sec consecutively (e.g., if the infant turned back to the center or looked at the caregiver, the floor, or the ceiling). If the infant turned away from the target by 30° in any direction for less than 2 sec and then turned back again, the trial went on, but the time spent looking away was not included in the orientation time. Thus, the maximum orientation time for a given trial was the duration of the entire sound file. The flashing red light remained on for the duration of the trial.

The familiarization phase consisted of alternating presentations of two short sequences until orientation times to each of them was at least 20 sec. On a given trial, repetitions of only one of the two sequences were played. The two sequences had the same words (either Set 1, "Leafy vegetables taste so good," or Set 2, "Rabbits eat leafy vegetables") and differed in prosody. One was a well-formed unit, the other was the corresponding prosodically ill-formed sequence. The side of the loudspeaker from which the stimuli were presented varied randomly from trial to trial. When the familiarization phase was over, infants heard eight test trials, four of each long passage presented in a random order. For both groups of infants, one test passage contained the familiarized well-formed unit, whereas the other test passage contained the familiarized ill-formed sequence. The materials were constructed so that the test passage containing the well-formed unit and the test passage containing the ill-formed sequence were reversed between the two groups.

An observer hidden behind the center panel looked through a peephole and recorded the direction and duration of the infant's headturns using a response box. The observer was not informed as to the group (familiarized with Set 1 or Set 2) to which the infant was assigned. Moreover, both the observer and the infant's caregiver wore earplugs and listened to masking music over tight-fitting closed headphones.

Results

Mean orientation times to the two passages heard during the test phase were calculated for each infant. Twenty-four of the 36 infants had longer orientation times for the passages containing the well-formed unit. Across all infants, the average orientation times were 13.33 sec (*SD* = 4.43) for the passages containing the well-formed unit and 11.30 sec (*SD* = 4.73) for the passages containing the ill-formed sequence. A paired *t* test indicated that this difference was significant, $t(35) = 2.58$, $p = .014$, 95% CI = $0.43 < 2.03 < 3.62$. Therefore, the results indicate that the infants listened significantly longer to the passage that contained the prosodically well-formed clause from the familiarization period than to the one that contained the ill-formed sequence.

Might this effect be due to infants' simply being more familiar with the well-formed sequence than the ill-formed sequence? Because the infants continued to hear both sequences during familiarization until at least 20 sec of listening time had accrued to each of them, it was possible that they could have heard the well-formed sequence significantly more often than the ill-formed sequence. However, this possibility can be ruled out. Mean familiarization time was 30.27 sec (*SD* = 6.38) to the well-formed sequence and 28.83 sec (*SD* = 6.24) to the ill-formed sequence. The difference was not reliable, $t(35) = 1.39$, $p > .10$.

Discussion

These results show that infants familiarized with a prosodically well-formed clause and a prosodically ill-formed sequence of the same words later recognized the well-formed clause better than the ill-formed sequence when each was embedded in a continuous passage of speech. This conclusion follows from the finding that, after familiarization, infants preferred to listen to the passage containing the well-formed unit over the passage containing the ill-formed sequence. A number of studies with infants of about the same age have shown that infants prefer to listen to passages in which previously familiarized units (e.g., single words) are embedded (Jusczyk & Aslin, 1995; Jusczyk, Hohne, & Bauman, 1999; Jusczyk, Houston, & Newsome, 1999). Our findings apply to word sequences as opposed to single words. Importantly, they also demonstrate a more subtle phenomenon: Given equal opportunities to become familiar with a well-formed and an ill-formed sequence, the former is better recognized than the latter when both of them later occur embedded in a (multiclause) passage. This effect occurs despite the fact that the ill-formed sequence, no less than the well-formed sequence, constitutes an acoustic match to a part of the passage.

The earlier results of Mandel and colleagues (Mandel et al., 1994; Mandel et al., 1996) supplied evidence that words spoken as a prosodically well-formed clause afforded 2-month-olds a memory advantage. How do our results with 6-month-olds allow us to go beyond these earlier findings? These results show that, by 6 months of age, a well-formed clause is better recognized even when its reoccurrence is embedded in a longer passage of continuous speech. Because infants will often hear continuous speech outside the laboratory, these findings justify a wider application to conditions under which language acquisition actually occurs.

Furthermore, our results raise the particularly interesting possibility that prosody influences how infants parse the test passages. Infants may be using well-formed prosodic envelopes to package the continuous passage into smaller units corresponding to its component clauses. Thus, during online processing of the passages, prosodic cues could favor the greater extractability of the embedded sequence of words that is well-formed over the embedded sequence that is ill-formed. On this account, infants favor the passage from which they can more easily extract an exemplar of the familiarization sequence. If this interpretation is correct, these findings constitute the first evidence that infants actually use prosodic cues in the online segmentation of continuous speech.

However, there is an alternative, not mutually exclusive way to account for the findings. This alternative makes two assumptions. One is that the well-formed sequence from familiarization is better remembered than the ill-formed sequence. This is plausible in light of previous findings (Mandel et al., 1994; Mandel et al., 1996), but it is not sufficient by itself to explain our observations here. In this case, the well-formed and ill-formed sequences constituted exactly the same series of words. Even if the well-formed familiarization sequence were more memorable, it should correspond to a matching series of words in both test passages. Thus, one has to add a second assumption, namely that the advantage for the passage containing the well-formed sequence was due to the fact that it contained an exact acoustic match (not just a word-by-word match) to the supposedly more memorable familiarization sequence.

Given the implications for language acquisition of the prosodic parsing account, we were interested in creating a situation in which the preference for the passage containing the well-formed clause could only be accounted for by prosodic parsing. One way to do so is to remove the acoustic match between the familiarization sequence and its manifestation in the test passage. Indeed, the alternative account relies crucially on the existence of an acoustic match between the sequences presented during familiarization and test, although the prosodic parsing account predicts that the effect observed in this experiment should also be obtained in the absence of an acoustic match. Consequently, the next two experiments investigated whether a preference for the passage containing the well-formed sequence could be obtained in the absence of an acoustic match between its realizations at the time of familiarization and test.

EXPERIMENT 2

A very strong test of the prosodic parsing account is to pit the potential advantage of prosodic-parsing cues in the absence of an acoustic match against the potential advantage of an acoustic match in the absence of effective parsing cues. If only one test passage has prosodic markers that highlight the well-formed sequence but only the other test passage includes an acoustic match to an ill-formed familiarization sequence, will infants still listen longer to the former? To do so, infants would have to more easily detect a nonacoustic match to a well-formed prosodic unit than an acoustic match to a prosodically ill-formed sequence. Such a result would dramatically favor the prosodic-parsing explanation. In contrast, the opposite finding, one favoring the passage with the ill-formed acoustic match, would strongly support the view that acoustic matching plays a central role in driving infant preferences. In this case, an acoustic match would be favored even when the sequences are not prosodically well-formed.

Method

Participants. Forty-eight American infants (21 girls, 27 boys) from monolingual English-speaking families participated in this experiment. The infants had an average age of 26 weeks, 5 days (range = 23 weeks, 4 days to 27 weeks, 0 days), or approximately 6 months. The data from an additional 8 infants were not included for the following reasons: becoming fussy and crying (4), experimental error (2), average orientation times shorter than 3 sec (1), and not turning to the lights (1).

Stimuli. Once again, the stimuli consisted of two test passages and four short sequences. A new female speaker of American English recorded the stimuli. The test passages were the same as those in Experiment 1. Acoustic analyses of the test passages by the new speaker indicated that the portions corresponding to the well-formed and ill-formed sequences differed in much the same manner as in Experiment 1. Silent gaps of 204 and 639 msec for Passages I and II, respectively, were present between the second and third words of the ill-formed sequences (cf. 60 and 132 msec for the same locations in the well-formed sequences). Also, the ill-formed sequences had a sharp rise in pitch and rms amplitude between the second and third words of the sequence, whereas the well-formed sequences showed no sign of either a pitch rise or an increase in rms amplitude in this location. Instead, for the latter sequences, pitch peaks occurred on the initial word of each sentence and on the stressed syllable of the word receiving emphatic stress (*vegetables* in Passage I and *good* in Passage II). In contrast to the previous experiment, there was no evidence of a systematic change in duration of the stressed syllable at the sentential boundary for the ill-formed sequences. In fact, the durations of the comparable syllables were actually longer for the well-formed sequences (perhaps an indication of the speaker's marking of the subject–predicate phrase boundary).

In contrast to Experiment 1, there were two types of familiarization sequences. For the prosodically ill-formed sequences, each sequence was excised from one of the test passages, just as in Experiment 1. By comparison, the prosodically well-formed sequences were excised from an entirely new set of passages to ensure that there would be no acoustic match between the word sequences presented during the familiarization and the test phases. The new well-formed sequences were excerpted from the following carrier passages:

(Ia) Guess what kind of leafy things taste so good? **Leafy vegetables taste so good.**

(IIa) I don't know anything about what rabbits drink. **Rabbits eat leafy vegetables.**

Notably, the excerpted sequences (in bold) were spoken so as to produce emphatic stress on a different syllable than for the comparable sequence in the test passages. Thus, the overall intonation contours of the well-formed sequences in familiarization and within the test passages differed from each other. Acoustic measurements verified that this was the case. Thus, secondary pitch peaks, as well as rms amplitude increases and longer syllable durations for the well-formed sequences used in familiarization occurred on *vegetables* and *leafy* for Sequences Ia and IIa, respectively, as opposed to *good* and *vegetables* for the comparable test passages.

Consequently, in this experiment, there was an acoustic match between the prosodically ill-formed familiarization sequences and their test counterparts but no exact acoustic match between the prosodically well-formed familiarization sequences and their test counterparts. The average duration of each new sequence was 2.28 sec. The average duration of the test passages was 6.55 sec. Once again, each sound file was made up of several repetitions of the short sequences or passages (five and three repetitions, respectively). The maximum trial duration of the familiarization sequences was 14.14 sec, whereas the maximum trial duration of the test passages was 21.20 sec.

Design. Infants were randomly assigned to one of two groups defined in terms of the set of sequences (Set 1 vs. Set 2) that they heard during the familiarization phase. There were 24 infants in each group.

Apparatus and procedure. These were the same as in Experiment 1, except that a Macintosh Quadra 650 replaced the PDP 11/73.

Results

Mean orientation times to the two passages heard during the test phase were calculated for each infant. Twenty-four of the 48 infants had longer orientation times for the passages containing the well-formed unit. Across all infants, the average orientation times were 9.18 sec (*SD* = 3.39) for the passages containing the well-formed unit and 8.78 sec (*SD* = 3.69) for the passages containing the ill-formed sequence.[1] A paired *t* test indicated that the difference between the mean listening times for the passages with the well-formed and ill-formed sequences was not significant, $t(47) = 0.76$, $p > .45$, 95% CI = –0.65 < 0.39 < 1.44. Thus, in contrast to Experiment 1, the infants in this experiment gave no evidence of detecting the well-formed sequence better than the ill-formed sequence within the passages.

Inspection of the amount of time infants spent listening to the well-formed and the ill-formed sequences during the familiarization phase revealed the unexpected finding that orientation times were longer for the ill-formed sequences. Mean familiarization time was 25.82 sec (*SD* = 4.26) for the well-formed sequences and 27.45 sec (*SD* = 2.81) for the ill-formed sequences, $t(47) = 2.19$, $p < .05$. Still, al-

[1]Orientation times in this experiment were much shorter overall than those in Experiment 1. There are a number of possible reasons for this, including the use of different speakers in Experiment 1 versus Experiments 2 and 3, the shorter duration of the test passages in Experiment 2 and 3 than in Experiment 1, and differences in the population of participants in the three experiments (Buffalo for Experiment 1 and Baltimore for Experiments 2 and 3). Such fluctuations in overall listening times are not uncommon in our experience in testing infants with this procedure, even when the materials and the population are the same. Consequently, the critical comparisons concern differences in infants' performances between conditions of the same experiment.

though the difference was reliable, it was small; it was less than 2 sec, about 6% of the total familiarization time for each sequence.

Discussion

This experiment was intended to provide a strong test of the prosodic parsing account in a situation in which either correct prosodic parsing of the test passages or acoustic identity could lead to the better detection of a familiarized sequence embedded in longer passages. No clear difference emerged in the infants' listening responses to the test passages, only a nonsignificant tendency in favor of the passage containing the prosodically well-formed sequence. Thus, removing the acoustic match between the well-formed familiarization sequence and its counterpart in the test passage, while leaving in place an acoustic match between the ill-formed sequence and its counterpart, did appear to affect infants' performance. In particular, they no longer showed a clear preference for the passage with the well-formed sequence. At the same time, it is evident that the availability of an acoustic match for the passage containing the prosodically ill-formed sequence was not sufficient to induce a preference for that passage.

One possible interpretation of these findings is that the pattern of results obtained in Experiment 1 was affected by both factors. That is, infants' better recognition of the well-formed embedding may have been influenced both by an acoustic match to the more memorable sequence (the well-formed sequence presented during familiarization) and by the greater ease of extracting the well-formed unit from the test passages on the basis of prosodic parsing cues. If the latter factor played a role, eliminating acoustic matches to both well-formed and ill-formed sequences in the test passages may reinstate the preference for the passage containing the well-formed sequence, the pattern that was observed in Experiment 1. The next experiment addressed this possibility.

Before this experiment is described, it is also relevant to mention a pilot study we conducted for other purposes but that permitted an assessment of the role of acoustic matching by itself. Six-month-olds were either familiarized with a single well-formed sequence or a single ill-formed sequence, then tested for their listening times for three different passages. One of these passages contained an exact acoustic match to the familiarization sequence, and one of them contained the same sequence in its alternative form (ill-formed at test if well-formed in familiarization or well-formed at test if ill-formed in familiarization). The third passage, irrelevant for current purposes, contained many of the same words as the familiarization sequence, but they were spoken in a different syntactic frame and, hence, a different order. Using some of the data from this pilot study, we can ask whether the acoustic match that was exclusive to the ill-formed sequence in Experiment 2 might have enhanced its detectability and, thus, canceled a competing advantage based on parsing cues for the well-formed sequence. For this purpose, we examined the listening times to test passages that contained the ill-formed version of the familiarization sequence. Specifically, we compared infants familiarized with that sequence (infants who could detect an exact acoustic match within the passage) to those familiarized with the well-formed version of the sequence (for whom no acoustic match was available within the passage). These means were 11.85 and 8.01 sec, respectively, $t(34) = 2.77$, $p < .01$, indicating better recognition of the ill-formed but acoustically matching familiarization sequence. In light of such evidence that acoustic matching to the ill-formed sequence may have been a competing effect in Experiment 2, we turn to Experiment 3, in which we eliminated the possibility of acoustic matching to both the well-formed and ill-formed sequences. Such a design should provide a more sensitive test for the role of prosodic parsing.

EXPERIMENT 3

Because our studies were the first to test infants for their recognition of clauses embedded within continuous speech, they were also the first that might implicate infants' use of prosodic information to parse the speech stream into clausal units. As we suggested, prosodic parsing could contribute to the recognizability of the well-formed units by allowing infants to more easily detect those units than the ill-formed sequences when they recur embedded in the test passages. If infants use prosody to parse their online representation of continuous speech into a string of clausal "packages," the well-formed sequence should cohere as an embedded unit more than the ill-formed sequence. As a result, infants should be better able to recognize that they have heard that sequence of words previously when it is a prosodically well-formed unit at its later hearing.

The results of the first two experiments are consistent with this prediction, but neither provides convincing evidence for it. The preference for the well-formed embedding in Experiment 1 can be explained alternatively as due to the acoustic match of that embedding to the better encoded familiarization sequence. The lack of any preference in Experiment 2 fails to support a strong and unambiguous interpretation of the results of Experiment 1 as due to prosodic-parsing effects. However, the findings of Experiment 2 may not be indicative of an absence of prosodic parsing. Rather, acoustic matching of the ill-formed familiarization and test sequences may have countered the emergence of a preference for the passage with the well-formed sequence. The latter possibility was eliminated in the design of Experiment 3. The materials for the new experiment were created to eliminate acoustic matches between either of the familiarized sequences and their embeddings in the test passages. Finding a preference for the passage with the well-formed sequence rather than the ill-formed sequence in this experiment would allow us to eliminate the alternative account of the results of Experiment 1 and provide unambiguous support for the involvement of prosodic parsing in our observations of sequence recognition.

Experiment 3 was also motivated by another issue. If infants only recognize a well-formed embedding of a previously encountered, well-formed short utterance when they match each other acoustically as in Experiment 1, the kind of ability we have been investigating would be of limited importance in language acquisition. Infants rarely, if ever, encounter such exact acoustic replicas in their everyday experience of language. Consequently, it is important to determine whether well-formed embedded units are better recognized than ill-formed embedded sequences when acoustic variation renders acoustic matching impossible. Had the results of Experiment 2 demonstrated a preference for the passage containing the well-formed sequence when no acoustic match was provided, the issue would have been resolved already. However, the failure to find such a preference in Experiment 2 is ambiguous for the reasons already discussed. Because acoustic matches were totally eliminated in Experiment 3, if infants showed better recognition of the well-formed embedded unit than the ill-formed embedded sequence, this would suggest that they could detect recurring clauses, an ability that could be useful in acquiring language.

Method

Participants. Thirty-six American infants (17 boys, 19 girls) from monolingual English-speaking families were tested. The infants had an average age of 25 weeks, 6 days (range = 23 weeks, 5 days to 28 weeks, 0 days), or approximately 6 months. Seven additional infants were tested but not included for the following reasons: becoming fussy and crying (6) and equipment failure (1).

Stimuli. Once again, the stimuli consisted of two test passages and four familiarization sequences. The recordings of the two test passages were identical to the ones used in Experiment 2. The prosodically well-formed familiarization sequences were also the same as in Experiment 2. New materials, produced by the same speaker, were used to create the prosodically ill-formed sequences. Specifically, these sequences were extracted from the following passages:

(Ib) Do rabbits eat green *leafy vegetables? Taste so good* is rarely encountered.
(IIb) What do big, strong *rabbits eat? Leafy vegetables* taste so good.

Note that, as for the well-formed sequences, the prosodically ill-formed sequences in these carrier passages were acoustically different from their counterparts in the test passages. The primary differences between the versions in familiarization and test were manifested in the pitch and amplitude characteristics of the samples. For Sequence Ib, in comparison to test passages, the familiarization sequences had pitch peaks and marked increases in rms amplitude on *vegetables.* For Sequence IIb, there was also evidence of significant pitch and amplitude changes for the sequences in familiarization and test. However, these occurred on the word *rabbit,* which had higher pitch and great rms amplitude in the familiarization sequence. Silent pauses at the internal sentence boundary were longer for the familiarization sequences (464 msec for Sequence Ib and 706 msec for Sequence IIb) as compared to the test passages pauses (204 and 639 msec, respectively). There were no systematic differences evident with respect to syllable durations of the ill-formed sequences used in familiarization and test. Still, ill-formed sequences in familiarization and in the test passages differed from each other acoustically in the other ways we have indicated.

The average durations of the short sequences and the passages, as well as the maximum familiarization and testing trial durations, were identical to those of Experiment 2.

Design. As in the two previous experiments, for purposes of counterbalancing, infants were randomly assigned to one of two groups defined in terms of the set of segments (Set 1 vs. Set 2) that they heard during the familiarization phase. There were 18 infants in each group.

Apparatus and procedure. These were the same as in Experiment 2.

Results

Mean orientation times to the two passages heard during the test phase were calculated for each infant. Twenty-seven of the 36 infants had longer orientation times for the passages containing the well-formed unit. Across all subjects, the average orientation times were 10.76 sec (*SD* = 4.46) for the passages containing the well-formed unit and 8.43 sec (*SD* = 3.63) for the passages containing the ill-formed sequence. A paired *t* test indicated that this difference was significant, $t(35) = 4.29$, $p = .0001$, 95% CI = 1.23 < 2.33 < 3.43. Hence, the infants listened longer to the passage containing the prosodically well-formed clausal unit, even though that unit was acoustically different from the sequences heard during familiarization.

There was no hint in this experiment that infants were differentially familiarized with the well-formed and ill-formed sequences. On average, they listened to the well-formed sequence during familiarization for 25.82 sec (*SD* = 4.04) and to the ill-formed sequence for 25.88 sec (*SD* = 3.68), $t(35) < 1$.

Discussion

The results of this experiment bear on two important issues regarding infants' processing of spoken language, prosodic parsing and clause recognition, both of which have implications for language acquisition. With regard to the latter, the findings establish that when an isolated sequence of words recurs as a prosodically well-formed clause within a passage of continuous speech, infants recognize it, and importantly, recognition does not depend on there being an acoustic match between its earlier and later occurrences. Thus, infants acquiring language are not limited to recognizing the reoccurrence of acoustically identical clausal exemplars, an event of very limited frequency in the normal language environment. Rather, even when the same words occur in different prosodic packages, as long as both of them are spoken with the prosody of a well-formed clause, infants appear to be able to detect their equivalence. Furthermore, recognition occurs even when the recurring unit is embedded in a continuous string of clauses.

Our results also provide the strongest indication so far that infants are not only sensitive to prosodic cues to clausal units but that they actually use those cues in on-line processing of speech to parse the linguistic input into clauses. Because neither the well-formed nor the ill-formed familiarization sequence constituted an acoustic match to a sequence embedded in the test passages, the better recognizability of the well-formed embedding suggests its better extractability from or unitizing within continuous speech. However, it is worth considering another possibility. Although neither the ill-formed nor the well-formed sequences were exact matches to the sequences in the test passages, might the well-formed sequences have been more acoustically similar to each other than were the ill-formed sequences? To investigate this, we used a curve fitting procedure to determine whether the well-formed versions of the sequences were more highly correlated with each other with respect to pitch, amplitude, and syllable durations than were the ill-formed versions of the sequences. Thus, there were six such comparisons, one each for the pitch, amplitude, and syllable durations of Sequences I and II. Of these comparisons, three indicated no significant differences between the well-formed and ill-formed sequences, two indicated a significant difference suggesting that the well-formed sequence was more similar, and one yielded a significant result suggesting that the ill-formed sequence was more similar. Thus, the lack of a consistent pattern here provides little support for the notion that the well-formed versions of the sequences were more acoustically similar than were the ill-formed versions.

What remains is the clear possibility that infants use prosody to bracket the input in a way that allows them to recognize a previously familiarized sequence that is embedded as a well-formed unit in a longer passage. Whereas the results of Experiment 1 can also be explained by an alternative account that appeals to the better memorability of the familiarization sequence that acoustically matches the well-formed embedding, any alternative account of the new results along these lines is far less plausible. That 6-month-old infants actually use prosodic cues to parse incoming speech is the more compelling conclusion. Of course, it is also the more newsworthy for accounts of language acquisition.

It is also interesting that the effect size in this experiment was, if anything, larger than the effect size in Experiment 1. The squared point biserial correlation measure described in Cohen (1988), r_{pb}^2, provides a good indication of the proportion of the variance accounted for by the independent variable. Although the r_{pb}^2 effects for both experiments are large according to Cohen's criteria (i.e., $r_{pb}^2 >$ .14), the proportion of the variance accounted for by the independent variable in Experiment 3 ($r_{pb}^2 = .34$) was twice as great as that in Experiment 1 ($r_{pb}^2 = .16$). Thus, eliminating the acoustic match between the familiarization sequences and the embedded parts of the test passages certainly did not reduce and may actually have enhanced the greater ease of detecting the recurrence of the well-formed sequence. It is possible that the presence of an acoustic match to the ill-formed sequence may have competed with the effect of prosodic parsing in Experiment 1.

GENERAL DISCUSSION

This series of experiments provided further information about how the grouping of words into prosodic units helps 6-month-old infants to process speech in a way that is useful for acquiring language. Our findings show that infants use prosodic marking to help them recognize the reoccurrence of a clause when it is embedded in a longer passage of speech. Infants who were familiarized with a prosodically well-formed sequence of words listened reliably longer to passages with this sequence embedded in them than they did to comparable passages containing familiar but prosodically ill-formed sequences. Furthermore, although the familiarized sequences used in Experiment 1 were exact acoustic matches to the material embedded in the test passages, this was not the case for Experiment 3. Nevertheless, infants displayed the same pattern of results in both experiments, suggesting that recognition of a short utterance is facilitated by prosodic well-formedness even when there are acoustic differences between its earlier and later occurrences.

Why are prosodically well-formed embedded clauses better recognized than the same sequences of words when they are not spoken as a prosodic unit? It appears likely that the better recognition of the well-formed than the ill-formed embedded sequence is due, at least in part, to the way that the multiclause passage is processed at the time of test. In particular, given prosodic-parsing cues, the well-formed unit is more easily extractible or unitized within its passage than the ill-formed unit is within its passage. Accordingly, the well-formed unit in the test passage is more easily recognizable as the familiarized sequence of words than is the ill-formed unit. Thus, the advantage of well-formedness, particularly as it appears convincingly in Experiment 3, constitutes the first suggestion in the literature that infants actually use prosodic cues in their online segmentation of continuous speech into syntactic units such as clauses.

By showing that prosody has an effect on how infants actually process normal speech, our findings add to the earlier results of Mandel and colleagues (Mandel et al., 1994; Mandel et al., 1996) in buttressing the case for a role for prosodic bootstrapping in language acquisition. Three conditions have to hold for the prosodic bootstrapping hypothesis to be viable. In the introduction, we noted that evidence already exists in favor of the first two conditions, namely (a) the existence of a fit between syntactic units and prosodic units marked consistently in the signal and (b) infants' sensitivity to those markers. This study focused on the third condition, namely infants' actual use of such prosodic information in speech processing. Our findings show that prosodic packaging of units corresponding to syntactic clauses helps infants recognize familiar sequences of words when they later appear embedded in continuous passages. Furthermore, the results suggest that when infants hear continuous passages, they use prosodic cues that mark clausal structure to parse the input as they process it.

The fact that language learners apparently rely on prosodic processing will not in itself guarantee the acquisition of the syntactic structure of the native language. As noted earlier, prosodic structures do not consistently pick out the same syntactic structures (Nespor & Vogel, 1986; Selkirk, 1984). However, a number of diverse theories of syntactic acquisition (Bates & MacWhinney, 1989; Berwick, 1985; Pinker, 1984; Wexler & Culicover, 1980) are in general agreement about two conditions that must hold for the acquisition of syntax. First, infants must have the ability to recognize and process major syntactic units (e.g., clauses and phrases). Second, infants must be able to segment speech into strings of words. Our studies build on those of Mandel and her colleagues (Mandel et al., 1994; Mandel et al., 1996). Together, these studies provide evidence for the existence of the first element by the time infants are 6 months old: Short strings of words spoken with clausal prosody are memorable, and they are recognizable as units during online processing of continuous speech. Experiments on early word segmentation provide evidence that infants at around the same age also have the second of the required abilities. A number of studies have now shown that word segmentation abilities develop at 7½ months of age and become more efficient during the second half of the first year of life (Echols, Crowhurst, & Childers, 1997; Jusczyk, 1997; Jusczyk & Aslin, 1995; Jusczyk, Hohne, & Bauman, 1999; Jusczyk, Houston, & Newsome, 1999; Morgan, 1996; Saffran, Aslin, & Newport, 1996). Therefore, it appears that in the second half of their first year, infants possess the rudimentary tools for grouping and segmenting speech that are required by current theories of syntax acquisition.

One other point that bears mention concerns infants' capacities for coping with speech variability (Eimas & Miller, 1980; Jusczyk, Pisoni, & Mullennix, 1992; Kuhl, 1979, 1983). In discrimination tasks, many studies show that infants as young as newborns can discriminate between lists of stimuli in a context of great acoustic variability (Bertoncini, Floccia, Nazzi, & Mehler, 1995; Bijeljac-Babic, Bertoncini, & Mehler, 1993; Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Floccia, & Bertoncini, 1998). With respect to recognition processes, evidence exists that 7½-month-olds can recognize familiarized words in fluent passages, even in the presence of acoustic differences (Jusczyk & Aslin, 1995; Jusczyk, Hohne, & Bauman, 1999; Jusczyk, Houston, & Newsome, 1999). These studies suggest that, from birth, infants can build representations of speech that generalize to new instances with different acoustic characteristics. Our findings fit nicely into this picture, adding to it the demonstration that infants are also able to cope with acoustic variability when recognizing speech fragments longer than single words. They show that infants can undertake a one-to-one matching of two acoustically different word sequences constituting well-formed prosodic units. This also suggests that infants do not necessarily need to be presented with many acoustically different instances of a stimulus type to be able to match two different exemplars.

It is of some interest to ask whether the way prosody enhances the recognizability of clauses within speech passages has any parallels in other domains. Indeed, a number of similarities have been observed in how young infants process speech and musical stimuli (Aslin, Jusczyk, & Pisoni, 1997). For example, in processing musical sequences, infants focus on relations among notes (i.e., melodic contour) rather than absolute pitch (Trehub, Bull, & Thorpe, 1984), much the same way that they do in generalizing across different speakers' voices (Kuhl, 1979). Similarly, infants are sensitive to the sequential reordering of notes in music (McCall & Melson, 1970; Melson & McCall, 1970; Trehub et al., 1984; Trehub, Thorpe, & Morrongiello, 1985), just as they are to the sequential ordering of different syllables in speech (Mandel et al., 1996). Eight-month-olds also have some ability to extract highly probable, repeated sequences of tones from a 2-min-long tone stream (Saffran, Johnson, Aslin, & Newport, 1999). Of even more specific relevance, it has been argued that groupings of phrases exist in music that are similar to the clausal units found in speech (Deutsch & Feroe, 1981; Lehrdahl & Jackendoff, 1983). Just as infants prefer speech passages in which artificially inserted pauses coincide with clause and phrase boundaries to ones in which pauses interrupt such units (Hirsh-Pasek et al., 1987; Jusczyk, Hirsh-Pasek, et al., 1992), so do they prefer the same type of marking of musical phrases in passages (Jusczyk & Krumhansl, 1993; Krumhansl & Jusczyk, 1990).

Consequently, we constructed a set of musical stimuli with parallels to the types of speech stimuli used in Experiment 1. In particular, we created a pair of musical passages so that each passage contained a well-formed musical sequence (i.e., it was a coherent musical phrase) that appeared as an ill-formed sequence in the other passage (i.e., it occurred as fragments of two adjoining phrases). The materials are notated in Figure 1. If 6-month-old infants use musical phrase structure in encoding and processing musical sequences, then, in parallel to Experiment 1, they

FIGURE 1 The passages used in the music study reported in the General Discussion. Well-formed sequences are underscored with a broad bar and ill-formed sequences with a narrow line.

should listen longer to the test passage that includes the well-formed sequence than to the one that includes the ill-formed sequence, after being familiarized with both sequences.

The average orientation times were 11.95 sec (*SD* = 4.28) for the passages containing the well-formed unit and 10.76 sec (*SD* = 3.98) for the passages containing the ill-formed sequence. Although the difference was only marginally significant, $t(35) = 1.97$, $p = .056$, 95% CI = $-0.34 < 1.19 < 2.42$, it is in the same direction as we observed with the speech stimuli in Experiment 1. The effect size for this music study was $r_{pb}^2 = .10$, which is in the medium range according to Cohen (1988). The latter value is smaller than for either Experiment 1 ($r_{pb}^2 = .16$) or Experiment 3 ($r_{pb}^2 = .34$). Hence, although the structural grouping effect that prosody affords for speech may have some parallels in music, any such grouping effect was weaker with the particular musical stimuli that we used. We note that each of the phrases within our musical stimuli is perhaps more similar to a clause within a multiclausal sentence than to a clause that stands alone as a sentence (as in our speech stimuli). It remains to be seen whether the same underlying mechanisms are involved in infants' processing and encoding of such highly structured materials as speech and music. Still, there is the suggestion that 6-month-olds' abilities to take advantage of inherent structure in encoding and processing information may not be limited to language.

In closing, let us return to the results of the main experiments. Together, they provide additional support for the view that young infants' sensitivity to prosodic structure plays a significant role in their online processing of speech. When the prosodic packaging of information embedded in passages was consistent with a previously heard sequence, that is, when the sequence was spoken as a prosodically well-formed clause within the passage, 6-month-olds were more apt to detect it. Moreover, infants' ability to detect such correspondences in speech was found not to depend on the existence of an exact acoustic match between the familiarized sequence and its counterpart in the passages. These findings add to the previous literature by showing that infants are not only sensitive to the prosodic organization of utterances but that they also make use of this organization in encoding and processing speech.

ACKNOWLEDGMENTS

This research was supported by a grant from the Fondation Fyssen to Thierry Nazzi, a sabbatical leave award from Swarthmore College to Deborah G. Kemler Nelson, and a Research Grant from the National Institute of Child Health and Human Development (15795) and a Senior Scientist Award from the National Institute of Mental Health (01490) to Peter W. Jusczyk.

We thank Eileen Crowley and Amy Gambon-Dennis for their help in recruiting and testing the infants; Denise Mandel, Sabrina Voelpel, and Theresa Allocco for help in recording the stimuli; Larry Nelson for composing the musical sequences used in the music study; Katherine Nuckel for help in conducting the acoustic analyses of the stimuli; Delphine Dahan for helpful comments on a previous version of this article; and Paul Luce and Trish van Zandt for their advice concerning some of the statistical analyses presented in the article.

REFERENCES

Aslin, R. N., Jusczyk, P. W., & Pisoni, D. B. (1997). Speech and auditory processing during infancy. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, & language* (5th ed., pp. 147–198). New York: Wiley.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–73). New York: Cambridge University Press.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue-trading relations. *Journal of Memory and Language, 30,* 644–663.

Beckman, M., & Edwards, J. (1990). Lengthening and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology: I. Between the grammar and physics of speech* (pp. 152–178). Cambridge, England: Cambridge University Press.

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother–child speech. *Phonetics, 14,* 303–309.

Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: Rhythmical basis of speech representations in neonates. *Language and Speech, 38,* 311–329.

Berwick, R. (1985). *The acquisition of syntactic knowledge.* Cambridge, MA: MIT Press.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do four-day-old infants categorize multisyllabic utterances? *Developmental Psychology, 29,* 711–721.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88,* 503–522.

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

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science, 209,* 1140–1141.

Ferreira, F. (1993). The creation of prosodic structure during sentence production. *Psychological Review, 100,* 233–253.

Fisher, C. (in press). From form to meaning: A role for structural analogy in the acquisition of language. In H. W. Reese (Ed.), *Advances in child development and behavior.* New York: Academic.

Fisher, C. L., & Tokura, H. (1996a). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development, 67,* 3192–3218.

Fisher, C. L., & Tokura, H. (1996b). Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 343–364). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gee, J., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15,* 411–458.

Gerken, L. A., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: Nine-month-olds' sensitivity to phonological vs syntactic phrases. *Cognition, 51,* 237–265.

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

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.

Jusczyk, P. W. (1989, April). *Perception of cues to clausal units in native and non-native languages.* Paper presented at the biennial meeting of the Society for Research in Child Development, Kansas City, MO.

Jusczyk, P. W. (1997). Finding and remembering words. *Current Directions in Psychological Science, 6,* 170–174.

Jusczyk, P. W. (1998a). Dividing and conquering linguistic input. In M. C. Gruber, K. Olson, & T. Wysocki (Eds.), *CLS 34: The panels* (Vol. 2). Chicago: University of Chicago.

Jusczyk, P. W. (1998b). Using the headturn preference procedure to study language acquisition. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (Vol. 12, pp. 188–204). Stamford, CT: Ablex.

Jusczyk, P. W. (1999). Narrowing the distance to language: One step at a time. *Journal of Communication Disorders, 32,* 207–222.

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

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.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 62,* 1465–1476.

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

Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 627–640.

Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month old infants. *Cognition, 43,* 253–291.

Kemler Nelson, D. G. (1989, April). *Developmental trends in infants' sensitivity to prosodic cues correlated with linguistic units.* Paper presented at the biennial meeting of the Society for Research in Child Development, Kansas City, MO.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright Cassidy, K. (1989). How prosodic cues in motherese might assist language learning. *Journal of Child Language, 16,* 55–68.

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.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics, 3,* 129–140.

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1,* 70–73.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America, 66,* 1668–1679.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6,* 263–285.

Lederer, A., & Kelly, M. H. (1991). *Prosodic correlates to the adjunct/complement distinction in motherese* (Papers & Reports on Child Language Development, Vol. 30). Stanford, CA: Stanford University.

Lehiste, I., Olive, J. P., & Streeter, L. (1976). The role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America, 60,* 1199–1202.

Lehrdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentential prosody help infants to organize and remember speech information? *Cognition, 53,* 155–180.

Mandel, D. R., Kemler Nelson, D. G., & Jusczyk, P. W. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development, 11,* 181–196.

McCall, R. B., & Melson, W. H. (1970). Amount of familiarization and the response to auditory discrepancies. *Child Development, 41,* 861–869.

Melson, W. H., & McCall, R. B. (1970). Attentional responses of 5 month girls to discrepant auditory stimuli. *Child Development, 41,* 1159–1171.

Morgan, J. L. (1986). *From simple input to complex grammar.* Cambridge, MA: MIT Press.

Morgan, J. L. (1994). Converging measures of speech segmentation in preverbal infants. *Infant Behavior and Development, 17,* 387–400.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language, 35,* 666–688.

Morgan, J. L., Swingley, D., & Miritai, K. (1993, April). *Infants listen longer to speech with extraneous noises inserted at clause boundaries.* Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, LA.

Nakatani, L., & Dukes, K. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America, 62,* 714–719.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception & Performance, 24,* 756–766.

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development, 21,* 779–784.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology.* Dordrecht, The Netherlands: Foris.

Peters, A. (1983). *The units of language acquisition.* Cambridge, England: Cambridge University Press.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Polka, L., Jusczyk, P. W., & Rvachew, S. (1995). Methods for studying speech perception in infants and children. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 49–89). Timonium, MD: York.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90,* 2956–2970.

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

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

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior, 23,* 450–466.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

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

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1985). Infants' perception of melodies: Changes in a single tone. *Infant Behavior and Development, 8,* 213–223.

Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.