

How Do 4-Day-Old Infants Categorize Multisyllabic Utterances?

Ranka Bijeljic-Babic, Josiane Bertoncini, and Jacques Mehler

Three experiments, using the high-amplitude sucking procedure, tested whether 4-day-old infants discriminate multisyllabic utterances on the basis of number of syllables or number of phonemes. Experiment 1 showed that infants discriminate 2 large sets of phonetically variable utterances composed of 2- vs. 3-CV (consonant-vowel) syllables. Experiment 2 was run to assess whether infants discriminated the 2 sets on the basis of duration differences between the 2- and 3-CV stimuli. Results indicate that reducing the duration differences does not affect infants' discrimination. Finally, Experiment 3 investigated whether infants discriminate 4- vs. 6-phoneme bisyllabic utterances. The results provide no evidence that infants are sensitive to such a change in number of phonemic constituents. Although not decisive, these results appear to be congruent with the hypothesis that infants perceptually structure complex speech inputs.

Models designed to account for speech processing in adults do not agree about the necessity of a prelexical level of representation during word recognition (for a review, see Frauenfelder & Tyler, 1987; Klatt, 1989). Some of them suggest that language users map the acoustic waveform onto units stored in the lexicon. But how do infants cope with the speech signal and represent it before they have constituted a lexicon? Acquisition of a lexicon requires a prelexical level of representation that mediates between acoustic inputs and memory traces that will constitute the lexicon (see Mehler, Dupoux, & Segui, 1991, for a discussion of this point). Several sublexical units have been considered to describe infants' prelexical representations. A number of mutually nonexclusive candidates can be evoked in this context, namely, phonemes, syllables, demi-syllables, moras, and feet. Moreover, it is possible that the efficiency of these different units depends on the development of the speech system and on the language that the child is in the process of acquiring. A review of the literature indicates that the phoneme

and the syllable have been most frequently put forward as the units that organize speech inputs.

Some evidence suggests that the syllable may behave like a perceptual unit, both for adults and children (Mehler, 1981; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Savin & Bever, 1970; Treiman, 1983; Walley, Smith, & Jusczyk, 1986). Furthermore, Bertoncini and Mehler (1981) found that young infants discriminate changes in phoneme ordering within a well-formed consonant-vowel-consonant (CVC) syllable (e.g., tap vs. pat) but failed to respond to the same kind of change in a CCC sequence (pst vs. tsp). However, when the same CCC clusters were embedded in a vowel context (i.e., upstu vs. utspu), the infants again showed reliable discrimination. These results suggest that even for very young infants, syllables have a privileged perceptual status. Presumably, a contrast incorporated into a well-formed syllable is more distinct or perceptually more salient than a similar one that is embedded in a nonsyllabic context. However, even if one grants that infants react more reliably to contrasts in good syllabic patterns, there is no indication that they organize speech in terms of syllabic units rather than phonetic segments.

The discrimination capacities of infants have been extensively investigated over the last two decades. Results from these experiments can be taken as evidence "for the existence of prelinguistic perceptual categories that are assumed to provide the basis for the acquisition of the corresponding adult phonetic category" (Miller & Eimas, 1983, p. 138). Although Miller and Eimas (1979) stated that they had found evidence that infants perceive organization at both the syllabic and segmental levels, very few studies have focused critically on that question. Some recent work suggests rather that young infants do not represent sets of variable CVs in terms of their phonetic segments (Bertoncini, Bijeljic-Babic, Jusczyk, Kennedy, & Mehler, 1988; Jusczyk & Derrah, 1987). The present research attempts to investigate more directly how infants organize and represent multisyllabic utterances.

A few studies have explored infant processing of multisyllabic utterances. Results show that infants are able to perceive some phonetic contrasts in multisyllabic contexts (Jusczyk &

Ranka Bijeljic-Babic (now at Laboratoire de Psychologie du Langage, Université de Poitiers, Poitiers, France), Josiane Bertoncini, and Jacques Mehler, Laboratoire de Sciences Cognitives et Psycholinguistique, Centre National de la Recherche Scientifique-Ecole des Hautes Etudes en Sciences Sociales, Paris, France.

This research was supported by grants from Centre National d'Etude des Télécommunications (CNET; Convention No. 00 790 92 45 DIT) and from Human Frontier Science Program ("Contrasting Language Phonologies" project).

We thank Christel Sorin (of CNET) for her help in the preparation of the first version of the stimulus materials, Michel Zerbib (of Centre National de la Recherche Scientifique; CNRS) and Jean-Luc Aucouturier (of CNRS) for their technical assistance, and Anne Christophe for providing helpful suggestions in statistical matter. We are also grateful to the staff at the Baudelocque and Notre Dame de Bon Secours Maternités in Paris for their kind cooperation. Finally, we also thank Peter Jusczyk for very helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Josiane Bertoncini, Laboratoire de Sciences Cognitives et Psycholinguistique, 54 Boulevard Raspail, F-75270 Paris Cedex 06, France.

Thompson, 1978) and are sensitive to prosodic cues like peak intensity (Bull, Eilers, & Oller, 1984) and vowel duration (Eilers, Bull, Oller, & Lewis, 1984). Moreover, Karzon (1985) found that 1- to 4-month-old infants discriminated /ra/ from /la/ when the target syllables were presented in the medial position of trisyllabic utterances, such as *marana* and *malana*, but only when the syllables were "exaggerately" stressed. This suggests that infants can use prosodic information in multisyllabic utterances to discriminate syllable components. Goodsitt, Morse, Ver Hoeve, and Cowan (1984) asked whether infants trained to react to the occurrence of a single syllable (e.g., /ba/) could recognize it when it appeared in one of two contexts, in a fairly redundant one (e.g., *tibati*, or *titiiba*) or in a nonuniform one (e.g., *batiko* or *tibako*). Their results showed that 6½-month-old infants can extract a token syllable from either of these contexts, although their performance is reliably better with the redundant one. It should be noted, however, that most of the aforementioned studies have used synthetic monosyllables combined into strings with a between-syllables interval that ranges from 5 ms to 50 ms. On the whole, these stimuli do not incorporate either coarticulation cues or differences in speaking rate, pitch, stress, and so forth that exist in natural utterances. These results indicate a precocious ability to perceive some fine acoustic and phonetic information within a multisyllabic string. However, these studies do not provide any indication as to how complex speech inputs are represented by infants or on infants' capacity to extract structural regularities across different multisyllabic utterances.

The first two experiments are designed to evaluate whether infants categorize and discriminate two sets of natural multisyllabic utterances that differ in number of constituent syllables. Given that CV represents the most typical and universally used form of syllable, we used CV strings as stimuli, arranged in two sets, one of two-CVs and one of three-CVs. If infants can discriminate such multi-CV utterances in spite of wide variations in duration and phonemic composition, we would have an indication that they organize multisyllabic sequences as strings composed of syllabic constituents.

Experiment I

If syllables are a salient element of speech, infants should be able to detect a change in the number of syllables in multisyllabic utterances. In the first experiment, newborns were habituated to a wide set of two-CV (or three-CV) utterances and then were tested with a new set of three-CV items (or two-CV items, respectively). We used naturally produced utterances so as to include all the features, including coarticulatory cues and intonation, that characterize multisyllabic words in continuous speech.

Method

Stimuli. All the stimuli used in the experiment were either bisyllabic (consisting of two-CV syllables) or trisyllabic (three-CV syllables). The CVs included one French consonant selected from b, d, g, p, t, k, r, l, f, v, s, z, š, ž, m, n, and ŋ and one vowel selected from a, e, i, o, and u. None of the items included phoneme duplication: Tokens like *rijo*, *zuti*, *kepa/mazopu*, *kesopa*, and *rekiu* were accepted, but tokens like *reri* or

rekiu were not. In addition, the probability for each CV to occur in first, second, or final position was the same.

One hundred and fifty items were selected (75 two-CVs and 75 three-CVs). They were all uttered by a native French female speaker who recorded each item embedded within the same carrier sentence "C'est XXX ca?" (for instance, *C'est baku ca?*: phonetically, *se baku sa?*) in order to produce similar prosodic patterns. As is the rule in French, the syllable in final position carried stress. The two- and three-CV stimuli were extracted from the carrier utterance by using a resident speech editing system on a PDP 11/73 computer.

The durations of the two-CV items ranged from 455 ms to 659 ms ($M = 551$ ms, $SD = 42.7$), and those of the three-CV items ranged from 637 ms to 894 ms ($M = 750$ ms, $SD = 47.3$). Figure 1 shows the distribution of duration of the items used in Experiment I.

Four audiotapes were prepared, two for each list of two- and three-CV items. Each series of 75 items was recorded in random order in such a way that items similar in duration or phonetic composition do not appear consecutively. Each series of stimuli lasted about 2 min and was repeated 10 times to obtain 20 min of stimulation on each tape.

Subjects. Subjects were healthy, full-term newborns (gestational age ranged from 38 to 41 weeks; $M = 39.5$ weeks) recruited at the Baudelocque Maternity Hospital in Paris, France. They had suffered no complications during pregnancy and delivery, and they were classified as "normal" after neurological evaluations on their first and third days of life (Amiel-Tison, 1977). All subjects weighed more than 2,800 g at birth (M birth weight = 3301 g, $SD = 340$) and had 5-min Apgar scores of 10.

Eighty-six subjects were tested on the third or fourth day after delivery (M age = 3.75 days). Forty-one infants were excluded for the following reasons: falling asleep before the shift (20), ceasing to suck during the course of the experiment (8), crying (7), failing to achieve the habituation criterion within 15 min (5), or experimenter error (1). Forty-five infants (26 girls and 19 boys) completed the experimental session. They were randomly assigned to one of two preshift stimulus conditions (stimulus set presented during habituation: two-CV stimuli or three-CV stimuli) and to one of two experimental conditions, either the control (no stimulus change: *bi-bi* or *tri-tri*) or the experimental (stimulus change: *bi-tri* or *tri-bi*). Thus the experimental group ($N = 30$) received one kind of stimuli (two-CV or three-CV items) during habituation and then shifted to the other (three-CV or two-CV items, respectively) during the test. In the control condition ($N = 15$), subjects were habituated with two-CV or three-CV items and continued listening to the same set of stimuli during the test.

Procedure. Infants were tested individually in a sound-attenuated chamber. They were awakened about 2 hr after feeding and were manipulated until a quiet, active state was obtained. They were then placed in a reclining position in a special "baby bath chair" that prevents head movements. The pacifier was held by an adjustable mechanical arm to avoid any intervention by the experimenter during the experimental session.

The experimental procedure was similar to the high-amplitude sucking technique devised by Siqueland and De Lucia (1969). We adjusted the criterion for high-amplitude sucking for each infant to obtain a baseline rate of between 20 and 50 sucks per minute. Stimuli were then presented contingent on the rate of high-amplitude sucking responses. Criterion sucks resulted in the presentation of one of the 75 stimuli from the preshift set. Which of the items the infant heard was unpredictable, because the series of 75 stimuli had been prerecorded in random order. On the tape, a stimulus onset occurred every 1.5 s. Thus, the subject might not hear all the 75 stimuli during the first presentation of the set, unless he or she produced high-amplitude sucks without interruption for 2 min. However, although it was possible for an infant to hear the same stimulus several times during the habitua-

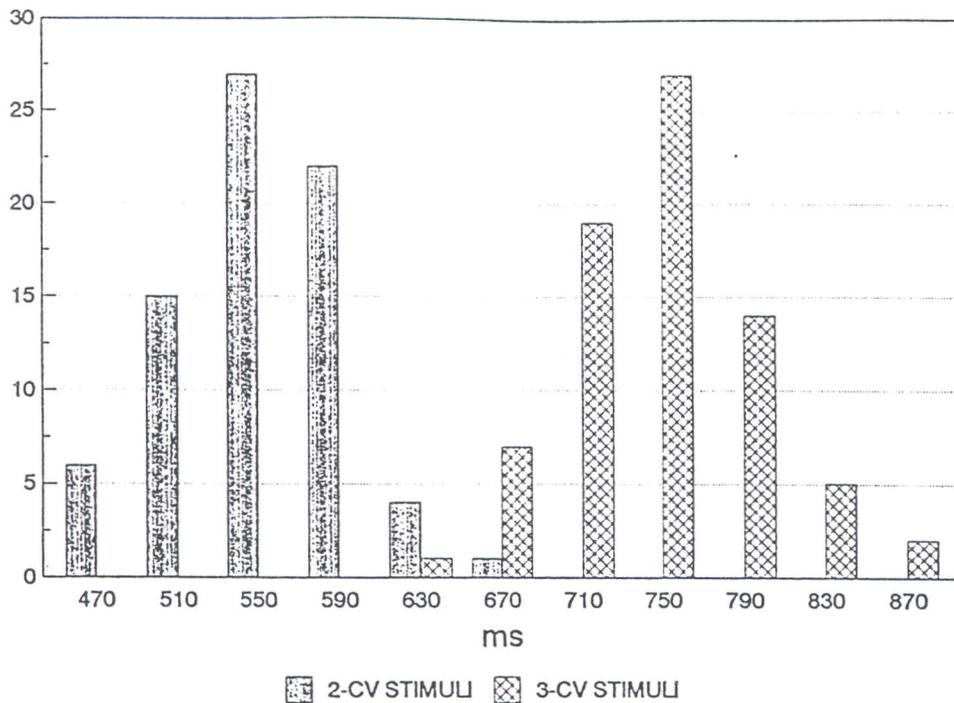


Figure 1. Distribution of durations of the two-CV (consonant-vowel) and three-CV stimuli used in Experiment 1. (Values correspond to the center of each 40-ms-wide category range.)

tion or the test phase, this only occurred with an interval of at least 112 s (75×1.5) between the two occurrences of the same item. Use of a programmable logic board ensured that all stimuli were uninterrupted during playback.

In summary, because the tape runs continuously with a stimulus ready to be played every 1.5 s, and given the temporal organization of nonnutritive sucking in bursts and pauses, we can consider that an infant who produces about 30 high-amplitude sucks per minute should receive about 25 different stimuli from the series during that minute.

The criterion for switching from the preshift to the postshift set of stimuli was a decrease in sucking rate of 25% or more over 2 consecutive minutes compared with the rate in the immediately preceding minute. Once criterion was reached, presentation shifted automatically and without interruption to the stimuli on the second tape recorder in both experimental and control conditions. The switch resulted in a change in the set of stimuli presented to the infants in the experimental condition and in the presentation of another recording of the same set of stimuli for the infants in the control condition. The conditions of stimulus presentation remain the same throughout the experiment. As a result, the items an infant heard just before the change and just after the change were completely unpredictable. The postshift period lasted 4 min.

Apparatus. We connected a pacifier to a Gould P23 pressure transducer and used an electronic custom-designed device to interface an IBM-PC compatible computer with the pressure transducer and the tape recorders. The multifunction electronic device detected the sucking responses and transformed them into digital output. It also ensured the uninterrupted presentation of each stimulus and controlled the channel switch electronically. The computer was programmed to deliver stimuli contingent on high-amplitude sucking. The items were delivered by two tape recorders (Tandberg TD 20A), a stereo amplifier (Scott 417 A), and a loudspeaker (Braun L 620).

Results

An overall analysis of variance (ANOVA) on sucking rates during the baseline, the 5 last preshift minutes, the 4 postshift minutes, and on habituation time (time to reach habituation criterion) was performed with group (experimental vs. control) and preshift stimulus condition (two-CV vs. three-CV stimulus sets) as main crossed factors.

There is no significant effect of group and preshift condition on any preshift measures. The baseline sucking rates were similar in the two groups, $F(1, 41) < 1$. There is no difference in the time to reach habituation criterion depending on the group (8.1 min for experimental group and 8.9 min for control group) or on the preshift condition (8.6 min for two-CV condition and 8.3 min for three-CV condition). Finally, the analysis of the sucking rates during the last 5 preshift minutes reveals no effect of group or preshift condition, $F(1, 41) < 1$ in each case.

To evaluate infant reaction to the change of stimuli, we compared the last 2 min of habituation and the first 2 min of test. These comparisons revealed a significant interaction between the shift (2 min before vs. 2 min after the change) and the group, $F(1, 41) = 8.824$, $p = .005$. This interaction is due to a significant increase in sucking rates in the experimental group, $F(1, 28) = 19.629$, $p = .0001$, and no difference in the control group, $F(1, 13) < 1$. This analysis revealed that the preshift stimulus condition did not interact with shift or with Group \times Shift interaction, $F(1, 41) < 1$ in each case. Post hoc comparisons revealed that each experimental group was significantly different from the control group with respect to the shift: *bi-tri*, $t(28) = 2.497$, $p < .01$, one-tailed; *tri-bi*, $t(28) = 2.763$, $p < .01$, one-

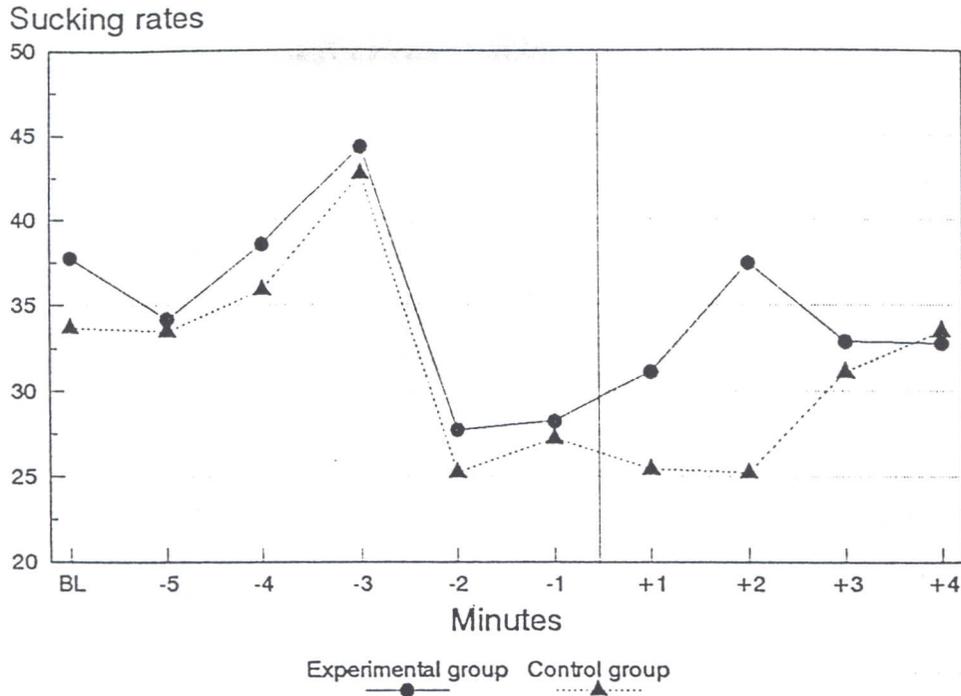


Figure 2. Sucking-rate averages during baseline (BL), last 5 min of habituation (-5 to -1), and 4 min of test (+1 to +4) for experimental and control groups in Experiment 1.

tailed. Figure 2 displays sucking-rate averages for experimental and control groups during the baseline, the last 5 min of habituation, and the 4 min of test.

Discussion

The results of this experiment indicate that 4-day-old neonates discriminated a shift from a set of two-CV items to a set of three-CV items, or vice-versa, over a broad range of phonetic variations. Indeed, infants noticed the difference between the two sets of multisyllabic utterances, regardless of the phonemic and durational diversity of the strings within each set. These discrimination responses may rely on infants' ability to extract a common perceptual pattern from each set of stimuli. One possible explanation is that they classify strings according to the number of their CV constituents, the only property maintained constant across the various items within each set. However, one should refrain from accepting this interpretation too hastily. Our experiment was designed to investigate whether infants can discriminate between two-CV and three-CV stimuli, when all characteristics of such utterances are congruent. The stimuli were naturally spoken items, in which duration was not equated. Thus, duration may have constituted the distinctive property between the sets of two-CV and three-CV utterances. In spite of a variation in duration within each set of stimuli, the three-CV items were consistently longer than the two-CV items. As can be seen in Figure 1, the distributions of the bisyllabic and trisyllabic item durations hardly do overlap (455–659 ms and 637–894 ms, respectively). Thus, one possible explanation for our results is that newborns classify utterances

according to their "typical" duration and are sensitive to the difference in overall durations between the two sets. Such an explanation, however, rests on the infants' hypothetical ability to notice the typical duration of multisyllables in one set of 75 different items, when the shorter items differ from the longer ones by 200–250 ms, and to react rapidly when they are presented with new items whose durations vary with the same ratios.

To our knowledge, no study has demonstrated that infants use time as a yardstick for establishing perceptual categories across phonetic variations. Moreover, studies on the effect of speaking rate on the perception of phonetic categories indicate that infant as well as adult listeners accommodate for changes in speaking rate by means of relational processes that cannot be described as being exclusively driven by temporal properties (Eimas & Miller, 1980; Miller & Volaitis, 1989). These results render implausible infant categorization of stimuli, or parts of stimuli, on the basis of their absolute durations. Nevertheless, before accepting such indirect evidence to settle the issue, we ran a second experiment to establish whether infants' ability to discriminate a change in number of syllables persists when durational differences are minimized. To achieve this, our second experiment used the same stimuli as those in Experiment 1, but the durations of two-CV and three-CV items were made to overlap in such a way that duration could no longer be a consistent cue on which infants might distinguish two-CVs from three-CVs.

If the infants base their discrimination mainly on the durational differences between the two sets, their performance should be affected when these differences are considerably re-

duced. On the other hand, if the infants discriminate two-CVs from three-CVs on the basis of nondurational properties, then they should replicate the same pattern of responses as in Experiment 1.

Experiment 2

Method

Stimuli. The same sets of stimuli as those in Experiment 1 were used in Experiment 2. To obtain overlapping distributions of durations, we submitted the two sets of stimuli to an algorithm designed to compress or extend speech signals. This was accomplished with a program developed by Centre National d'Etudes des Télécommunications, France (see Charpentier & Stella, 1986) that was implemented on a PDP 11/73.¹ Item durations were modified, but the global contour of the energy and the intelligibility of the syllables were spared. Every two-CV and three-CV item used in Experiment 1 was randomly assigned to a compression or extension factor.² The reason for applying both extension and compression procedures to two-CVs as well as to three-CVs was to ensure that each set of stimuli not be characterized by one speech rate (i.e., a rapid one for three-CVs and a slow one for two-CVs). Thus, to avoid having only compressed three-CV stimuli to be discriminated from only extended two-CV stimuli, we treated each stimulus with the compression/extension algorithm independently of its two- or three-CV structure and irrespective of its original duration. Obviously, applying such a procedure resulted in an enlarged duration variation within each set of stimuli. In that sense, duration no longer appeared as a property that could readily characterize each set of stimuli. Nevertheless, to minimize the duration difference between the two sets, we more frequently shortened than expanded the three-CV items and more frequently expanded than shortened the two-CV items. The resulting set of two-CV items included 27% compressed utterances (11% at 0.75 and 16% at 0.90) and 52% extended utterances (23% at 1.15 and 29% at 1.25) mixed with 21% unchanged sequences. Among the three-CV items, 55% were shortened (30% at 0.75 and 25% at 0.90), 24% were extended (15% at 1.15 and 9% at 1.25), and 21% remained unchanged. As can be seen in Figure 3, the resulting distributions were largely overlapping (overlap involved 75% of stimuli) when compared with those of the original stimuli, displayed in Figure 1. Duration of the two-CV items now ranged from 330 ms to 740 ms ($M = 582$ ms, $SD = 104.5$) and that of the three-CV items ranged from 510 ms to 960 ms ($M = 697$ ms, $SD = 114$). Although the duration distributions do not fully overlap, the difference between the duration averages is now less than 120 ms, and the dispersion of durations within each set is much larger, because the difference between the longer and the shorter items is more than 400 ms.

Subjects. Seventy infants were tested at the Notre Dame de Bon Secours Maternity, in Paris. They were selected according to the same criteria as those used in Experiment 1. Fifty infants (23 girls and 27 boys; M age = 4.7 days) completed the experiment. They were full-term newborns (M gestational age = 39.2 weeks; M birth weight = 3,372.1 g, $SD = 569$) with no detectable abnormality or deficit. Twenty subjects were excluded from the data analysis for the following reasons: falling asleep before the shift (8), crying (7), failing to achieve the habituation criterion within 15 min (1), and experimenter error (4).

Procedure. The procedure was identical to that described in the previous experiment, except that the amplitude threshold for high-amplitude sucking was fixed for all the subjects, and the baseline period lasted until the infant's sucking rate was stabilized between 20 and 50 sucks per minute. As in Experiment 1, the group factor (experimental vs. control) was crossed with the preshift stimulus condition factor (two-CV vs. three-CV). Thirty subjects were randomly assigned to the experimental condition and shifted from two-CV to three-CV stimuli

(or vice versa). Twenty subjects were included in the control condition and were habituated and tested with the same set of stimuli (two-CV or three-CV).

The stimuli were recorded under the same conditions as in Experiment 1. The 75 compressed and extended stimuli of each set were recorded in random order, with one stimulus occurring every 1.5 s. Randomization ensured that items whose durations or compression/extension rates were similar did not appear in a sequence.

Apparatus. A slightly different configuration of the audio equipment was used in the second experiment: two tape recorders (Tascam Porta 05), a stereo amplifier (Rotel RA-820BX3), and a loudspeaker (Martin DB 92).

Results

Overall ANOVAs were performed on baseline, the 5 last preshift minutes, the 4 postshift minutes, and habituation time. These analyses revealed no significant effect of the preshift stimulus condition, nor of the group factor on preshift measures. Preshift stimulus condition has no effect on habituation time, $F(1, 46) < 1$, nor on the sucking rates during the last 5 preshift minutes, $F(1, 46) < 1$. The experimental and control groups produced similar sucking rates in the baseline, $F(1, 46) < 1$, and during the 5 preshift minutes, $F(1, 46) = 1.820$. Moreover, time to reach the habituation criterion was similar, $F(1, 46) < 1$, for the experimental group (8.2 min) and the control group (7.9 min).

Comparisons between the last 2 min of habituation and the first 2 min of test were used to test the infants' reaction to the change of stimuli. These comparisons revealed a significant interaction between the shift (2 min before vs. 2 min after the change) and the group, $F(1, 46) = 4.159$, $p = .047$. This interaction is due to a significant increase in sucking rates by the experimental group, $F(1, 28) = 4.718$, $p = .038$, and no difference in the control group, $F(1, 18) < 1$. Neither the Preshift Condition \times Shift interaction nor the triple Preshift Condition \times Group \times Shift interaction was significant, $F(1, 46) < 1$ in each case.

Post hoc comparisons revealed that each experimental group was significantly different from the control group with respect to the shift: *bi-tri*, $t(33) = 2.05$, $p < .025$ (one-tailed); and *tri-bi*, $t(33) = 1.79$, $p < .05$ (one-tailed).

Figure 4 displays sucking-rate averages during the baseline, the last 5 min of habituation, and the 4 min of the test, for the experimental and control groups.

¹ Our compression algorithm uses the pitch synchronous overlap and add (PSOLA) technique. The algorithm basically takes successive chunks of a signal (whole periods for voiced segments) and averages them into a single chunk. The number of adjacent chunks to be averaged depends on the compression rate, and an adequate weighting function is chosen to ensure a perfectly smooth result. This technique does not alter spectral components (e.g., formants or pitch) of the signal. A description of the algorithm PSOLA is detailed in Dupoux and Mehler (1990).

² The maximum compression rates used in the present study (i.e., 0.75) only reduced the durations to 75% of their original lengths. These rates are much smaller than the ones used by Dupoux and Mehler (1990), who found that speech comprehension remained nearly perfect even when the signal was compressed by as much as 50%.

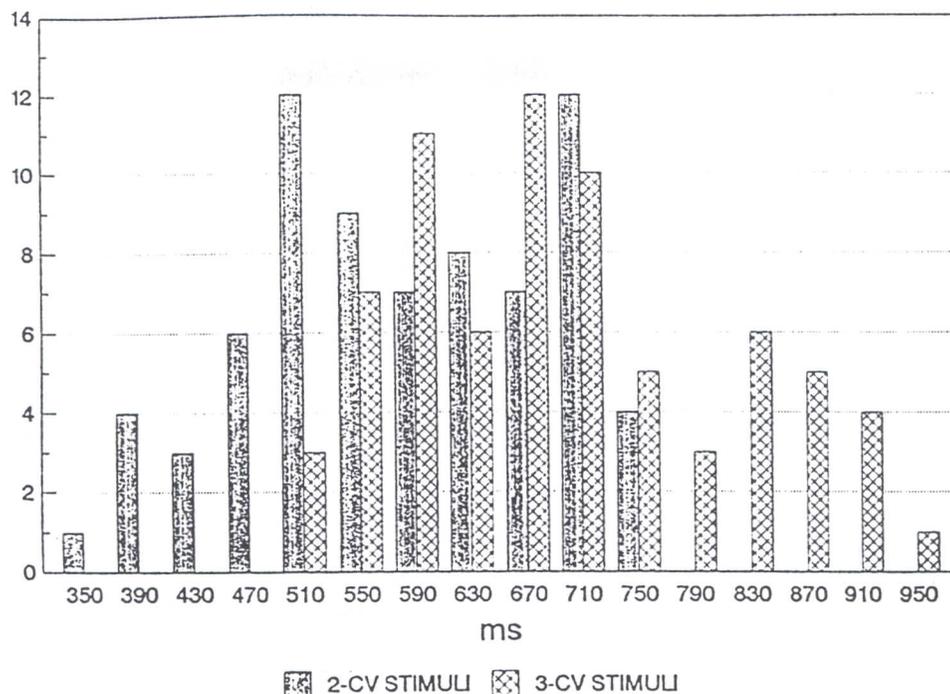


Figure 3. Distribution of durations of the compressed and extended stimuli used in Experiment 2. (CV = consonant-vowel.)

Discussion

The results of Experiment 2 corroborate those of Experiment 1. They confirm that 4-day-old infants discriminate two-CV from three-CV patterns. Furthermore, they suggest that the discrimination ability does not merely rely on the durational characteristics of the stimuli.

We designed Experiment 2 to verify whether infants continue to exhibit discrimination responses when durational differences between the two sets of stimuli are minimized. We chose to use exactly the same sets of stimuli and reduce, as much as possible, the difference between two-CV and three-CV durations. To maintain the "naturalness" of the stimuli, we applied low rates of compression and extension. In addition, to avoid introducing differences in speaking rate between two-CVs and three-CVs, we submitted the two sets to both extension and compression procedures. If we had treated distinctively the original three-CVs and two-CVs (i.e., if we had compressed the three-CVs and extended the two-CVs), we would have introduced a new confound based on speech rate differences between the two sets. But we know from studies on both speech and nonspeech stimuli that infants are sensitive to changes in rate (Eimas & Miller, 1980; Jusczyk, Pisoni, Reed, Fernald, & Myers, 1983). The resulting sets of stimuli obviously do not completely overlap, but the relationship between the two distributions is very different from what it is in Experiment 1, in terms of both average durations and amount of variability within each set.

If in Experiment 1 the infants have based their discrimination responses merely on the durational properties of the stim-

uli, it is conceivable that their performance would be altered when these properties were changed. On the contrary, the results obtained in Experiment 2 show that infants continue to exhibit a reliable reaction to the shift from two-CV to three-CV stimuli, similar to that observed in Experiment 1. Thus, their performance appears to be unaffected by a modification of the stimuli that makes duration cues less relevant.

However, the fact that the two distributions do not fully overlap means that, after the change, it is possible for the infants to receive few stimuli that fall outside the overlap (i.e., the two-CVs that are shorter or the three-CVs that are longer). Although we cannot radically exclude the possibility of an infant reacting to those outlying stimuli, it seems highly improbable. Once the habituation phase is completed, an infant may have heard most of the 75 items that range from 330 to 740 ms for the two-CVs and from 510 to 960 ms for the three-CVs. After the switch, the infant receives 25 new stimuli, among which at most 5 or 6 are shorter or longer than the stimuli he or she has heard before. How do 4-day-old neonates notice 5 items scattered through the 25 new stimuli, even though they all share structural properties in terms of number of syllables? Such an intricate capacity has never been reported in the infant speech perception literature. Moreover, in a study specially designed to test the infants' use of duration cues in speech perception, Eilers (1977), using 2- to 3-month-old infants, failed to observe discrimination between two single syllables that differed purely on overall duration.

Although the hypothesis that infants discriminate multisyllabic utterances solely on the basis of durational differences cannot be completely ruled out, it seems far less plausible than the alternative hypothesis that infants react to differences in

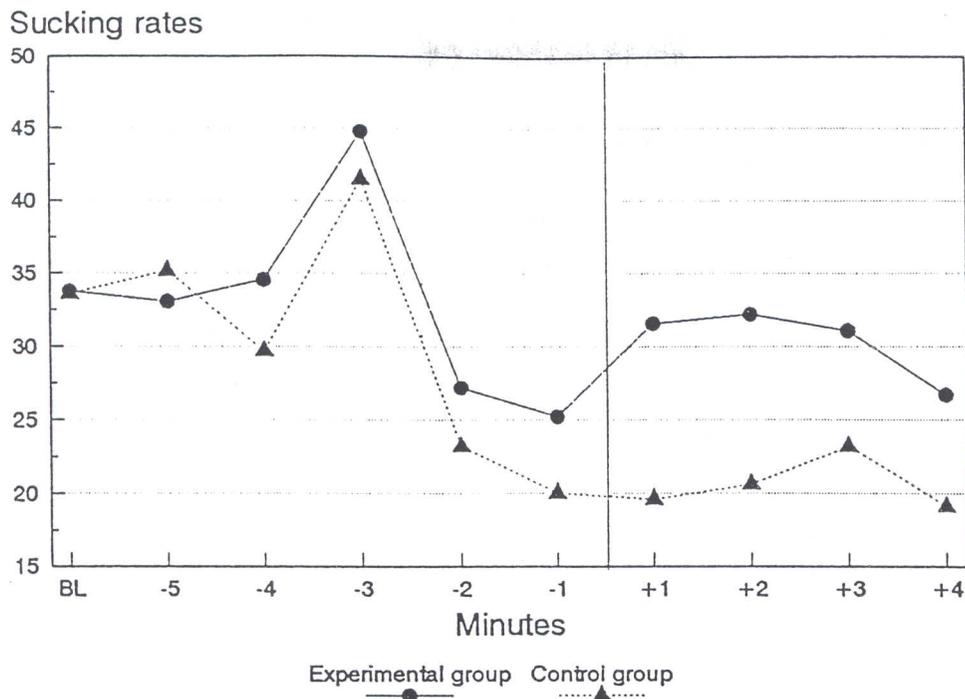


Figure 4. Sucking-rate averages during baseline (BL), last 5 min of habituation (-5 to -1), and 4 min of test (+1 to +4) for experimental and control groups in Experiment 2.

structural properties. Further experiments are needed to provide a decisive conclusion.

Taken as a whole, the results of Experiments 1 and 2 suggest that, even when faced with a large amount of stimulus variability, newborn infants still seem to be able to extract an invariant property characterizing each list of multisyllabic sequences. This property could be associated in some way with the number of CV constituents, which in our stimulus material corresponds to the number of syllabic units (two vs. three) as well as to the number of phonemic constituents (four vs. six). Thus, the infants may have categorized the multisyllabic sequences by organizing them either in terms of syllabic structure or in terms of phonemic constituents. In Experiment 3, we attempt to establish whether number of phonemes, irrespective of syllabic structure, is a parameter used by young infants in perceiving multisyllabic utterances.

Experiment 3

If infants are sensitive to the number of phonemes, they should discriminate stimuli that differ in number of phonemes but that are equated in terms of number of syllables. One way in which this can be tested is to contrast four- to six-phoneme items within bisyllabic or trisyllabic items. However, arranging four phonemes in trisyllabic strings leads one to include items like CVVV or VVVC, which are not very representative of trisyllabic utterances. Thus, we constructed two sets of varied four- versus six-phoneme utterances, all of which were composed of two syllables. This design was destined to allow us to

test whether infants discriminate four- versus six-phoneme patterns, irrespective of the number of syllabic constituents.

Method

Stimuli. We used two sets of 35 different bisyllabic utterances recorded by a French female speaker. Different types of syllable structures were equally distributed over the two sets of stimuli. The number of open and closed final syllables, the number of initial and final consonant clusters, and the distribution of the five vowels (a, e, i, o, u) were similar in both sets. Clusters were selected so as to conform to the phonotactic rules of French. One set consisted of 35 bisyllabic items, each composed of four phonemes. The structures used were CVCV (5), VCCV (5), CCVV (5), VCVC (10), CVVC (5), and VVCC (5). In the other set, the 35 bisyllabic items were composed of six phonemes. The structures were CVCCCV (7), CCVCCV (7), VCCCV (1), CVCCVC (4), CVCVCC (3), CCVCVC (3), CCVVCC (4), VCCVC (3), and VCCVCC (3).³ The constraint of having only bisyllabic items introduced structural differences between the sets (e.g., consonant clusters are more frequent in the six-phoneme set than in the four-phoneme set). In addition, sequences with three consecutive consonants (e.g., *alprim*) or four (e.g., *abstro*) appeared only in the six-phoneme set. The average duration of the six-phoneme items was 736 ms (range = 541 to 937 ms), and that of the four-phoneme items was 656 ms (range = 492 to 844 ms).

Subjects. Sixty-five infants were tested at the Baudelocque Maternity Hospital in Paris. They were selected according to the same crite-

³ Examples of four-phoneme items include *rifu*, *iblo*, *gria*, *azur*, and *duer*. Examples of six-phoneme items include *suldri*, *treklu*, *kafest*, *alprim*, and *novibl*.

ria as that used in Experiments 1 and 2. Twenty-five subjects were excluded from the data analysis for the following reasons: falling asleep before the shift (9), crying or ceasing to suck (9), failing to achieve the habituation criterion within 15 min (3), and experimenter error or technical problems (4). Forty infants (13 girls and 27 boys; M age = 3.2 days) completed the experiment. They were full-term newborns (M gestational age = 39.6 weeks; M birth weight = 3318.1 g, SD = 343) with no detectable abnormality or deficit.

Procedure. The procedure and apparatus were identical to those described in Experiment 2. As in the previous two experiments, the group factor was crossed with the preshift stimulus condition. Forty subjects were randomly assigned to the experimental or control conditions (10 subjects each). The experimental group (20 subjects) shifted from four- to six-phoneme items or from six- to four-phoneme items. The 20 subjects in the control condition were habituated and tested with four-phoneme or six-phoneme items. The stimuli were delivered under the same conditions as in the two previous experiments.

Results

The data were analyzed as in the previous experiments. First, the overall ANOVA revealed no significant effect of the preshift condition or of the group factor on any preshift measures. There is no difference in the habituation time or in the sucking rates during the 5 preshift minutes between the two preshift conditions, $F(1, 36) < 1$ in each case. The experimental and the control groups displayed no differences in baseline sucking rates, $F(1, 36) < 1$, or in habituation time (8.3 min and 9.15 min, respectively), $F(1, 36) = 1.037$. The sucking rates during the 5 preshift minutes in the two groups did not differ significantly, $F(1, 36) < 1$. The sucking-rate averages for experimental and control groups are displayed in Figure 5.

The ANOVA indicated that the Group \times Shift interaction (2 min before vs. 2 min after the change) was not significant but was very close to the critical level of .05, $F(1, 36) = 3.803$, $p = .059$. Contrary to the pattern of results generally observed in the high-amplitude sucking procedure, this tendency is not due to an increase in responses by the experimental subjects, but rather to an unusual increase in the responses by the control group. Separate comparisons revealed that there was no effect of shift in the experimental group, $F(1, 18) = 0.875$, whereas for the control group there was a significant increase in sucking rate between the last 2 min of habituation and the first 2 min of test, $F(1, 18) = 12.525$, $p = .002$.

To explore whether this paradoxical pattern of results actually indicates that the experimental groups exhibit significantly smaller differences in sucking rate compared with those in the corresponding control groups, we made post hoc comparisons. Both comparisons between the experimental groups and the control groups did not reach the critical level of .05: when the four- versus six-phoneme condition was compared with the four- versus four-phoneme condition, $t(18) = 1.59$, $p < .10$ (one-tailed); and when the six- versus four-phoneme condition was compared with the six- versus six-phoneme condition, $t(18) = 1.16$, $p > .10$ (one-tailed). Nevertheless, the observed pattern of results remains surprising. Close inspection of the parameters that could have affected the responses in the control group (e.g., selection criterion or baseline sucking rate) did not provide any indication about a potential difference between that group and the experimental one. Although it is true that one often has to cope with high variability in the data in studies with very young infants, it is very uncommon to obtain a significant response

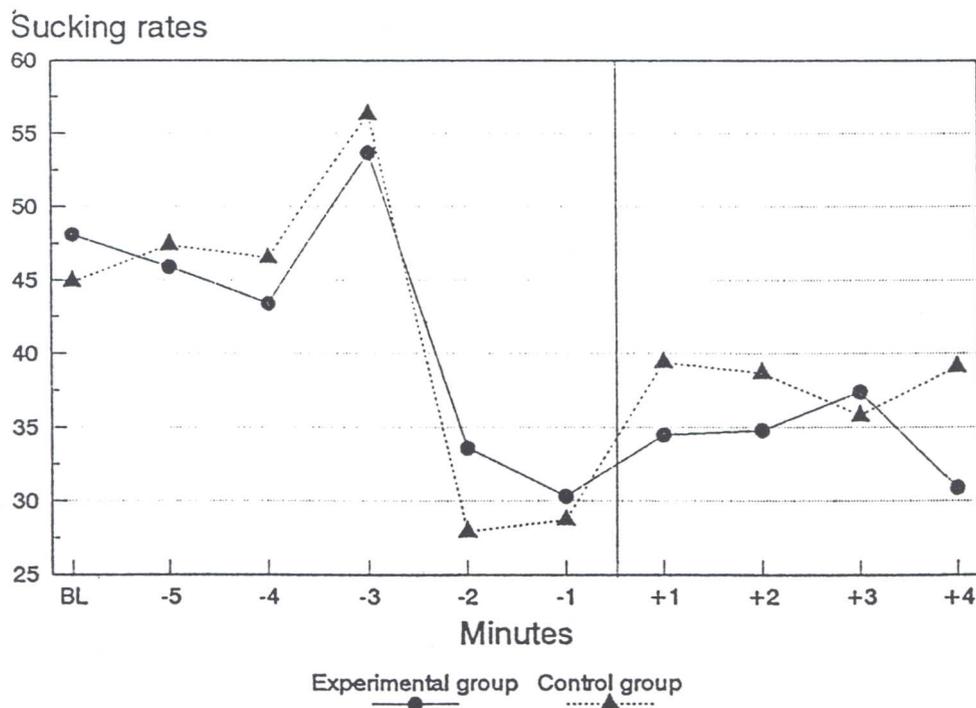


Figure 5. Sucking-rate averages during baseline (BL), last 5 min of habituation (-5 to -1), and 4 min of test (+1 to +4) for experimental and control groups in Experiment 3.

recovery in control groups as large as the one observed in Experiment 3. Our control subjects display a pattern of results that could statistically have occurred by chance in 1 case out of 20. We suspect that this incongruity can only be attributed to random error. To further investigate this peculiar outcome, we applied a regression analysis to estimate how significantly some critical variables contribute to the variability of postshift sucking rates (see Appendix).

Discussion

The third experiment was specifically designed to verify whether the discrimination responses observed in the previous two experiments might have arisen from the difference in number of phonemes (four vs. six) in the stimulus sets used in Experiments 1 and 2. By systematically varying the syllabic structure of bisyllabic items, we were able to test discrimination between sets of four- and six-phoneme items.

The absence of any significant recovery in sucking in the experimental conditions might suggest that a change from four- to six-phonemes is not in itself a difference to which neonates react. Moreover, experimental subjects in Experiment 3 also failed to detect duration differences that are of the same order as in Experiment 2. The higher proportion of complex structures (i.e., consonant clusters) in the six-phoneme set does not appear to be sufficient for the infants to discriminate the two kinds of stimuli. However, the peculiar behavior of the control group in Experiment 3 cautions us in drawing conclusions. At the very least, these results provide no indication that the infants in Experiments 1 and 2 could have categorized and discriminated multisyllabic utterances on the basis of their phonemic structures. Unfortunately, although these results are congruous with the view that infants do not organize complex speech stimuli as strings of phonemic units, they cannot be taken as robust evidence in favor of a syllabic level of organization. Further experiments are needed to investigate precisely whether infants structure speech inputs preferentially at the syllabic or at the segmental level, when these two levels are systematically decorrelated in stimulus materials.

General Discussion

Some studies have shown that infants are able to detect a change in the number of items in small collections of visual stimuli (Antell & Keating, 1983; Starkey & Cooper, 1980; Strauss & Curtis, 1981; Treiber & Wilcox, 1984). The fact that infants are able to use numerosity to categorize displays indicates that they perceive objects in their visual environment as distinct entities. The perception of numerosity has not been explored in the auditory modality (but see Moore, Benenson, Reznick, Peterson, & Kagan, 1987; Starkey, Spelke, & Gelman, 1983). However, if infants show similar capacities in discriminating auditory stimuli, this may be an indication that they perceive such stimuli as strings of auditory "discrete" events. The present study is an attempt to check which unit is used by infants to represent multisyllabic utterances.

Many previous experiments have investigated infants' ability to distinguish phonetic contrasts in isolated syllables, generally using an invariant vowel context (for a review, see Jusczyk,

1985). Our experiment, on the other hand, focused on infants' ability to process highly variable sets of naturally spoken multisyllabic utterances. Within each set of 75 items, our stimuli had varied durations and phonetic composition but shared the same number of CV constituents. Can infants extract this invariant property to form distinct categories of multisyllabic utterances? We assume that their success in discriminating two- from three-CV sets depends mainly on their ability to form a representation of the items related to number of CVs. Before accepting this view, we examined to what extent other putative properties like duration and number of segments could be involved in infants' processing of multisyllabic utterances.

First, we questioned whether the relationship between number of CVs and duration is the major property on which infants in our experiments based their responses. In Experiment 1, this relationship was very consistent, because all of the two-CVs were shorter than the three-CVs; in Experiment 2, most durations of two-CVs and three-CVs varied within the same range. In spite of this difference, the results show that infants react similarly to the change in number of CVs in both cases. In addition, in Experiment 3, the infants failed to discriminate the two sets of stimuli, notwithstanding the fact that the durational differences were of the same order as in Experiment 2. Taken together, these data do not support the idea that infants discriminate multisyllabic stimuli merely on the basis of overall duration. Second, we tested the infants' ability to discriminate bisyllabic patterns on number of phonemes. Using varied syllabic structures, we failed to observe discrimination between four- and six-phoneme sequences. Thus, the number of phonemic constituents per se does not appear to be within the scope of infants' discrimination abilities when they are asked to process highly variable multisyllabic utterances. The present data suggest, however, that infants can extract some organizational property from large sets of complex and variable speech stimuli.

In conclusion, our study explores the possibility that neonates organize multisyllabic utterances on the basis of some structural dimension. One possible interpretation of our data is that neonates use information related to number of syllabic constituents. But the fundamental question about the level of organization used by infants to represent the continuous speech signal must be addressed in future research if we are to know whether syllables (or some related parameter, i.e., vowel nuclei) are the primary units used by infants to initiate speech processing. Such structural information would act as a precursor leading the perceptual system to attend to the appropriate units for segmenting and representing continuous speech. In this context, several important questions should be raised. In particular, is the primary unit of representation universal, and how does it fit in with a segmentation process that might be language specific (Cutler, Mehler, Norris, & Segui, 1983, 1986)?

References

- Amiel-Tison, C. (1977). Standardizing the physical examination during the first year. In L. Gluck (Ed.), *Intrauterine asphyxia and the developing fetal brain* (pp. 473-485). Chicago: Year Book Medical.
- Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, 54, 695-706.

- Bertoncini, J., Bijeljic-Babic, R., Jusczyk, P. W., Kennedy, L., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, *117*, 21-33.
- Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development*, *4*, 247-260.
- Bull, D., Eilers, R. E., & Oller, D. K. (1984). Infants' discrimination of intensity variation in multisyllabic stimuli. *Journal of the Acoustical Society of America*, *76*, 13-17.
- Charpentier, F. J., & Stella, M. G. (1986). Diphone synthesis using an overlap-add technique for speech waveform concatenation. *Proceedings of the International Conference ASSP Tokyo* (pp. 2015-2018).
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature*, *304*, 159-160.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*, 385-400.
- Dupoux, E., & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language*, *29*, 316-335.
- Eilers, R. E. (1977). Context sensitive perception of naturally produced stops and fricative consonants by infants. *Journal of the Acoustical Society of America*, *61*, 1321-1336.
- Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *Journal of the Acoustical Society of America*, *75*, 1213-1218.
- Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science*, *209*, 1140-1141.
- Frauenfelder, U. H., & Tyler, L. K. (1987). *Spoken word recognition*. Cambridge, MA: MIT Press.
- Goodsitt, J. V., Morse, P. A., Ver Hoeve, J. N., & Cowan, N. (1984). Infant speech recognition in multisyllabic contexts. *Child Development*, *55*, 903-910.
- Jusczyk, P. W. (1985). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion* (pp. 199-229). Hillsdale, NJ: Erlbaum.
- Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology*, *23*, 648-654.
- Jusczyk, P. W., Pisoni, D. B., Reed, M., Fernald, A., & Myers, M. (1983). Infants' discrimination of the duration of a rapid spectrum change in nonspeech signals. *Science*, *222*, 175-177.
- Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics*, *23*, 105-109.
- Karzon, R. (1985). Discrimination of polysyllabic sequences by one- to four-month-old infants. *Journal of Experimental Child Psychology*, *39*, 326-342.
- Klatt, D. H. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169-226). Cambridge, MA: MIT Press.
- Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society; B* *295*, 333-352.
- Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*, 298-305.
- Mehler, J., Dupoux, E., & Segui, J. (1991). Constraining models of lexical access: The onset of word recognition. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 236-262). Cambridge, MA: Bradford Books.
- Miller, J. L., & Eimas, P. D. (1979). Organization in infant speech perception. *Canadian Journal of Psychology*, *33*, 353-367.
- Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition*, *13*, 135-165.
- Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*, 505-512.
- Moore, D., Benenson, J., Reznick, J. S., Peterson, M., & Kagan, J. (1987). Effect of auditory numerical information on infants' looking behavior: Contradictory evidence. *Developmental Psychology*, *23*, 665-670.
- Savin, H., & Bever, T. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, *9*, 295-302.
- Siqueland, E. R., & De Lucia, C. A. (1969). Visual reinforcement of nonnutritive sucking in human infants. *Science*, *165*, 1144-1146.
- Starkey, P., & Cooper, R. G., Jr. (1980). Perception of numbers by human infants. *Science*, *210*, 1033-1035.
- Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, *222*, 179-181.
- Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerosity. *Child Development*, *52*, 1146-1152.
- Treiber, F., & Wilcox, S. (1984). Discrimination of number by infants. *Infant Behavior and Development*, *7*, 93-100.
- Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, *15*, 49-74.
- Walley, A. C., Smith, L. B., & Jusczyk, P. W. (1986). The role of phonemes and syllables in the perceived similarity of speech sounds for children. *Memory & Cognition*, *14*, 220-229.

Appendix

Regression Analysis

A regression analysis allows one to test for any potential difference after a switch in stimulation while factoring out any difference before the switch. A standard regression analysis, including sucking rates before switch and dummy variables for the group factor, appears to be the proper statistical procedure in this case. In addition, it has the advantage of allowing us to use all 3 min before switch (instead of only the last 2 min as is the case in comparisons between the last 2 preshift minutes and the postshift ones). It also avoids unwarranted assumptions concerning the relationship between the sucking rates before and after switch: Indeed, computing a difference is equivalent to approximating by 1 the regression coefficient between before and after, instead of estimating its true value from the sucking rates themselves.

A regression analysis, using the mean sucking rates for 2 min after the switch as the dependent variable and the last 3 min before switch as independent variables, as well as two dummy variables for the two between-subjects factors already used in the ANOVAs (group: experimental vs. control; preshift condition: stimulus set presented during habituation) was performed on the results of Experiment 3. The group variable did not significantly contribute in the multiple regression, $t(34) = -1.5$, $p = .122$, nor did the preshift condition variable, $t(34) =$

-1.0 , $p = .31$. The results of the regression analysis clearly indicate that there is no significant difference between the postshift performance of the control and experimental groups in Experiment 3, when variability in preshift sucking rates is taken into account by factoring it out. A regression analysis was also performed on the results of Experiments 1 and 2. For Experiment 1, the preshift condition variable did not contribute to the multiple regression, $t(39) = 0.17$, $p = .8$, but the group variable contributed significantly, $t(39) = 3.24$, $p = .003$. Very similar results were obtained for Experiment 2. The group variable contributed significantly in the multiple regression, $t(44) = 2.83$, $p = .007$, and the preshift condition variable did not, $t(44) = -0.2$, $p = .8$. This analysis confirms that the group variable had a very significant effect only in Experiments 1 and 2 but did not reach a significant level in Experiment 3.

Received November 25, 1991

Revision received September 30, 1992

Accepted October 12, 1992 ■