

# Perception of syllable timing by prebabbling infants

Carol A. Fowler<sup>a)</sup> and Mary R. Smith<sup>b)</sup>  
*Haskins Laboratories, New Haven, Connecticut 06510*

Louis G. Tassinary<sup>c)</sup>  
*Dartmouth College, Hanover, New Hampshire 03755*

(Received 31 January 1985; accepted for publication 27 November 1985)

Adults hear alternating syllables with isochronous syllable onset-onset times as having a long-short, alternating rhythm when the syllables differ in initial consonant. This occurs because adults attend to syllable-internal events, called the "P centers" or "stress beats," rather than to syllable onsets. Thus they report that stress-beat aligned speech is isochronous and stress-beat aligned clicks are synchronized with the speech. The question asked here is whether, like adults, infants attend to the timing of syllable stress beats. In experiment 1, infants showed differences in time to habituate to sequences of alternating monosyllables, [bad] and [strad], having two different onset-onset times (onset- and stress-beat-timed) and two different placements of clicks on the syllables (on syllable onsets and on stress beats). Infants habituated more slowly to sequences with clicks on the stress beats than to sequences with clicks on syllable onsets and most slowly of all to stress-beat-timed speech with clicks on the stress beats. To interpret these findings, a second experiment was run using sequences only of the syllable [strad] so that speech timing measured according to onsets and stress beats was the same. Syllables had isochronous timing or a long-short alternating rhythm, corresponding to two possible ways of hearing the stress-beat-timed speech of experiment 1. In addition, two patterns of click placement were compared, uniform and syncopated, corresponding to two ways of hearing the stress-beat aligned clicks of experiment 1. The patterns of sucking times in the two experiments match exactly if stress-beat aligned speech in experiment 1 is identified with the isochronous speech of experiment 2 and the stress-beat aligned clicks of experiment 1 match with the uniformly timed clicks of experiment 2. It is inferred from this correspondence that infants perceive stress beats and stress-beat timing of syllables as adults do.

PACS numbers: 43.71.An, 43.71.Es

## INTRODUCTION

Our research examines perception of syllable timing by infants. Specifically, it is designed to ask whether, like adults, infants perceive the timing of syllables by reference to the syllables' "P centers," rather than to the timing of onsets of acoustic energy. To motivate the research, we first outline the relevant findings on adult perceivers and talkers.

When adult listeners judge the timing of sequences of different spoken words or nonsense syllables, they attend to different intervals than investigators have measured. Thus sequences of digits with isochronous intervals between measured onsets<sup>1</sup> do not sound isochronous to listeners (Morton *et al.*, 1976). Moreover, listeners who are given an opportunity to adjust alternating pairs of digits (e.g., "five ten five ten..."), until the sequence sounds isochronous, introduce systematic departures from isochrony as measured from the onsets of acoustic energy of successive digits (Morton *et al.*, 1976; see, also, Marcus, 1981). Morton *et al.* (1976) proposed that each spoken word has a "psychological moment of occurrence," called the "P center," that follows the onset of acoustic energy for a word by an interval determined by the word's acoustic composition. As yet, no acoustic marker

of the P center has been found, although markers have been sought in the amplitude envelopes and intonation contours of words (Morton *et al.*, 1976; Rapp, 1971; Tuller and Fowler, 1981) and in their segmental structure (Morton *et al.*, 1976; Marcus, 1981; Fowler and Tassinary, 1981).<sup>2</sup>

When talkers are asked to produce sequences of syllables in time with a real (Rapp, 1971; Fowler and Tassinary, 1981) or imaginary (Fowler, 1979; Tuller and Fowler, 1980) metronome pulse, they produce just the departures from isochrony that listeners require to hear the sequences as evenly timed. Accordingly, sequences of syllables produced in time to a metronome pulse offer a means of locating acoustic markers of the P center, if there are any, because the pulse enables corresponding points in different syllables to be identified. However, in sequences produced to a real metronome (Rapp, 1971; Fowler and Tassinary, 1981), the pulses coincide with no obvious acoustic marker in the syllable. Nevertheless, the metronome pulses may mark the P centers. Clicks placed in syllables where the metronome pulse tends to be located in productions of the same sequences are heard as synchronized with the syllables (Allen, 1972b, and see our pilot research below). Following Allen (1972a,b), we will call this location in the syllable the "stress beat"; we assume that it either coincides with the P center or is temporally equidistant from it across syllables. Placed elsewhere, for example, on onsets of acoustic energy for the syllables

<sup>a)</sup> Also at Dartmouth College, Hanover, New Hampshire 03755.

<sup>b)</sup> Now at MRC Applied Psychology Unit, Cambridge, England.

<sup>c)</sup> Now at University of Iowa, Iowa City, Iowa 52242.

(henceforth "syllable onsets"), clicks do not sound synchronized with most syllables.

The only isochronously timed aspects of a syllable that have yet been identified in productions such as those just described are articulatory. Tuller and Fowler (1980) found that muscle activity for each of the phonetic segments in sequences of rhyming CVC's was isochronous when talkers were instructed to produce isochronous syllables; this was the case even for sequences in which acoustic measures revealed marked departures from isochrony. For example, the sequence [bak fak bak...], produced by talkers in time to an imagined metronome, exhibited isochronous activity of the orbicularis oris muscle surrounding the lips. However, intervals from [bak] to [fak], as measured from onsets of acoustic energy for the syllables, are short relative to the complementary intervals. This discrepancy between the articulatory and acoustic measures is readily understood. Lip activity for [b] effects lip closure and, in these productions, a silent closure interval. The lip activity for the [f] causes the lower lip to approximate the upper teeth, giving rise to a noisy closure interval. Therefore, the talkers aligned a silent closure for the [b] with a noisy closure for the [f].

Two other characteristics of P-center-timed speech require mention to motivate our infancy research. First, in productions of syllables under isochrony instructions, intervals measured from onset of regular glottal pulsing for the vowel show smaller discrepancies than intervals measured from syllable onset; however, these intervals, too, show systematic departures from measured isochrony when muscle activity is isochronous. The departures are opposite in sign from those characterizing intervals between measured syllable onsets (Fowler and Tassinari, 1981). Second and concomitantly, in syllables with prevocalic consonants of relatively long measured duration, such as [s], the metronome pulse falls earlier relative to measured vowel onset than in syllables with shorter-duration prevocalic consonants, such as [b]. Moreover, there is essentially a millisecond-for-millisecond correspondence between this shift in the metronome location and the measured shortening of the vowel as the duration of prevocalic consonants increases (Fowler and Tassinari, 1981). We ascribe the measured shortening of the vowel to progressive coarticulatory covering over of the vowel by the prevocalic consonants in the syllable as they increase in articulatory and acoustic duration (Fowler, 1983, 1984) and we hypothesize that the stress beat corresponds to the isochronously produced vowel (Tuller and Fowler, 1980; Fowler and Tassinari, 1981).

In view of the finding that articulatory timing is isochronous in these productions, one interpretation of listeners' timing judgments and of their judgments that clicks located on stress beats are synchronized with syllable production is that they reflect detection of information for the talker's articulatory vowel timing in the acoustic speech signal (Fowler, 1979; Tuller and Fowler, 1980; Fowler and Tassinari, 1981). These acoustic consequences of vowel timing have not been identified, however, and Marcus (1981) has shown that the relative timing of different monosyllables in psychologically isochronous sequences can be predicted from acoustic-phonetic measures (the durations of any con-

sonants before the vowel and of the remainder of the syllable, differently weighted), with no explicit reference to articulatory timing.

For a variety of reasons, it is of interest to study the development of perception of syllable timing, however P centers are ultimately explained. Studies of a variety of languages suggest that attention to P-center timing is general to users of languages of different rhythmical types, including the stress-timed languages, English (e.g., Morton *et al.*, 1976) and Dutch (Eling *et al.*, 1980), the syllable-timed language, Spanish (Hoequist, 1983), and the mora-timed language, Japanese (Hoequist, 1983). This may imply that there are few degrees of freedom available to language users in their perception of syllable timing and to languages in their expressions of the different rhythms of languages. Correspondingly, it allows the possibility that infants will hear syllable timing as adults do, having had little or no experience with spoken language. An alternative possibility, particularly if it is the case that mature listeners extract acoustic information about articulatory timing, is that young listeners benefit from detecting their own speechlike articulations as they are reflected acoustically, and therefore develop attention to P-center timing only once they begin babbling, or even at some later phase of language development.

Our experiments investigate perception of syllable timing by infants averaging 4 months of age, asking whether it resembles that of adults even before infants begin to produce the speechlike articulations characteristic of "canonical" babbling (Oller, 1980) and well before they produce their first words. If it does, we can infer that infants need not hear the acoustic consequences of their speechlike babbling in order to perceive syllable timing as mature listeners do.

## I. EXPERIMENT 1

In the present studies, we used a non-nutritive sucking, habituation procedure (e.g., Eimas *et al.*, 1971) to ask whether infants hear syllable timing as adults do. To address this question, we presented infants with sequences of alternating monosyllables that are isochronous as measured acoustically (henceforth "acoustic-onset-aligned sequences" or simply "onset-aligned sequences")—and therefore sound to adults as if they have a long-short alternating rhythm—or with sequences having a long-short rhythm, as measured acoustically, that sound isochronous to adults (henceforth "stress-beat aligned"). In addition, we presented click sequences simultaneously with the speech. The clicks coincided either with the onsets of acoustic energy of the syllables or with the perceived stress beats of the syllables (as determined by pilot research on adults described below). According to the judgments of adults, the clicks sounded synchronized with the syllables only when they coincided with the stress beats.

Table I displays the four stimulus conditions in which the infants participated during the acquisition phase of the first experiment. The cells in the table show how the sequences are judged by adults.

Our research design has two parts. In experiment 1, we measure infants' responsiveness (operationalized as time-to-

TABLE I. The four acquisition-phase conditions of experiment 1, and their perception by adults.

		Click timing	
		Acoustic-onset aligned	Stress-beat aligned
Speech timing	Acoustic-onset aligned	Clicks are isochronous and they are heard as such; speech sounds anisochronous; speech and clicks sound nonsynchronous	Clicks are anisochronous and they are heard as such; speech sounds anisochronous; speech and clicks sound synchronous
	Stress-beat aligned	Clicks are anisochronous and they are heard as such; speech sounds isochronous; speech and clicks sound nonsynchronous	Clicks are isochronous and they are heard as such; speech sounds isochronous; speech and clicks sound synchronous

habituate) to the different sequences. Any differences we find will not be interpretable in terms of the infants' perception of syllable timing, however, until we compare the pattern of differences to a pattern of times to habituate to sequences of known perceived timing. We make this comparison in experiment 2. The purpose of experiment 1, then, is to discover whether a pattern of time-to-habituate differences can be obtained for the sequences of syllables and clicks whose perceived timing by infants we hope to discover. Experiment 2 was designed to interpret the pattern.

We had some grounds for expecting infants to show differential responsiveness to the sequences used in experiment 1. Kuhl (1976) found differences in time to habituate to various speech stimuli in a non-nutritive sucking procedure. When just one syllable was presented repeatedly, infants habituated in 7.76 min; time to habituate increased to 8.76, 10.32, and 11.26 min among infants hearing pairs of syllables differing, respectively, in pitch, in vowel, or in both dimensions. Kuhl suggests that time-to-habituate may reflect the time required for the infant to "learn" the stimulus set." (p. 276). A more general interpretation is that time to habituate reflects infants' interest in continuing to listen to stimulation, where interest may be sparked in a variety of ways.

Research by Demaney *et al.* (1977) shows that 2- to 3-month-old infants are sensitive to fairly subtle differences among auditorily presented rhythms when discrimination is indexed by recovery of looking at a geometrical form following habituation to one rhythmic sequence and a shift to another. Allen *et al.* (1977) have shown that 5- to 6-month-old infants recognize the equivalence of paired rhythms, one tonal and the other light flashes. Thus, if infants habituate to a sequence of tones and then are shifted to a light sequence (or vice versa), they show less recovery if the light and tonal sequences share a rhythmic pattern than if they do not. This study suggested that infants might detect the congruence or its lack between click and speech timing in our experiment.

Finally, several studies have shown that, in a head-turning paradigm measuring preference, 4- to 6-month-old infants will prefer an event congruent with a sound they are hearing to an incongruent event (e.g., Dodd, 1979; Kuhl and Meltzoff, 1982; MacKain *et al.*, 1983; Spelke, 1976, 1979). Moreover, they will do so when congruence and incongruence are determined entirely by the phase relations between the seen and heard events or entirely by their relative tempos

(Spelke, 1979). We considered it possible, then, that another influence on time to habituate, besides encoding difficulty as suggested by Kuhl (1976; also see Humphrey *et al.*, 1979), may be the perceived congruity of the speech and clicks.

We had no experimental grounds for expecting infants to habituate differentially to isochronous sequences and sequences with long-short alternating rhythms. However, infants might find the rhythms differentially interesting in either of two respects. First, whereas the isochronous rhythm is an ostensibly natural rhythm of spoken languages, the long-short rhythm is not. [That is, whereas isochronous intervals between stressed syllables is a very common underlying rhythm across languages, no languages have been found to alternate long and short intervals between stresses (e.g., Hayes, 1980; Prince, 1983).] Second, the long-short rhythm is more complex than the isochronous rhythm in that it includes two different time intervals, rather than just one. Following Kuhl's interpretation of her findings, we might expect infants to suck longer before habituating to the long-short rhythm.

In two respects, our use of the non-nutritive sucking, habituation procedure is somewhat novel and requires some comment. First, in most studies of discrimination of speech by infants, the infants' rate of sucking determines the rate at which syllables are presented (e.g., Eimas *et al.*, 1971). This was not possible in our experiments, in which timing was a manipulated variable. Accordingly, in our procedure, infants' sucking rates determined the amplitude at which the speech sequences were presented. Using an amplitude contingency, we find profiles of the infants' rates of sucking over the successive minutes of the experimental session similar to those obtained using the traditional contingency (see Figs. 2 and 4). Therefore, we have some confidence that the procedure can be used to examine infants' perception of syllable timing. A second unusual aspect of our procedure (but, see Kuhl, 1976) is our use of time-to-habituate as a dependent measure. Generally, in studies using the habituation procedure, researchers are interested in the infants' ability to discriminate acquisition and test speech stimuli. In our first experiment, too, we test for discrimination, and we look at recovery following a shift in syllable timing from acquisition to test. However, to ask the question of major interest to us—whether infants perceive stress-beat timed syllables as isochronous (and concomitantly, whether they hear clicks on

the stress beats as synchronized with the speech)—we need to look for group differences in responsiveness during acquisition. Time-to-habituate proved a measure sensitive to group differences.

## A. Methods

### 1. Design

Infants were assigned to one of the four experimental conditions or four control conditions on a rotating basis. (That is, in every set of eight infants run in succession, each experimental and control condition was represented once. This procedure was abandoned when the last few infants were run to make group sizes nearly equal.) In each condition, infants heard a sequence consisting of single tokens of two syllables, [bad] and [strad], repeated in alternation and presented with a sequence of clicks superimposed. We selected these syllables because they show large differences in stress beat location, according to adult users of the language (Fowler and Tassinary, 1981). In doing so, we hoped to maximize the likelihood that infants could discriminate the timing differences we employed and hence maximize our chances of observing differences in responsiveness to the different timing sequences. The experimental and control conditions used the different acquisition-phase sequences described in Table I. Timing of the speech and relative timing of the clicks that distinguished the acquisition-phase sequences were determined in pilot research on adults (reported in the Appendix). After habituation was detected in the acquisition phase (see our habituation criteria below), infants in the experimental conditions were shifted to a different one of the four sequences for the test phase. They were shifted to a sequence in which the absolute timing of the clicks remained as before, but the timing of the speech was changed from onset-aligned to stress-beat-aligned or vice versa. Each experimental condition had a matched control condition that shared its acquisition phase, but continued the same sequence during the test phase as in the acquisition phase. In all, then, there were three main independent variables: group (experimental or control), speech timing (onset-aligned, stress-beat-aligned), and relative click timing (coincident with syllable onset or with the stress beat). The main dependent measures were time-to-habituate and a discrimination measure. The measure of time-to-habituate compared the number of minutes of sucking in the acquisition phase (that is, until habituation) across groups of infants hearing the four different stimulus sequences. The discrimination measure compared the number of sucks in the last acquisition phase minute with the average number of sucks in the first 2 min after the shift point.

### 2. Stimulus materials

Sequences were isolated productions of the syllables [bad] (308 ms) and [strad] (474 ms) produced by the first author.<sup>3</sup> Just one token of each syllable was used; they are displayed in Fig. 1. The syllables were digitized at a 20-kHz sampling rate and low-pass filtered at 10 kHz. They were presented under computer control, in one of two different timing rhythms. Onset-aligned productions had a 600 ms

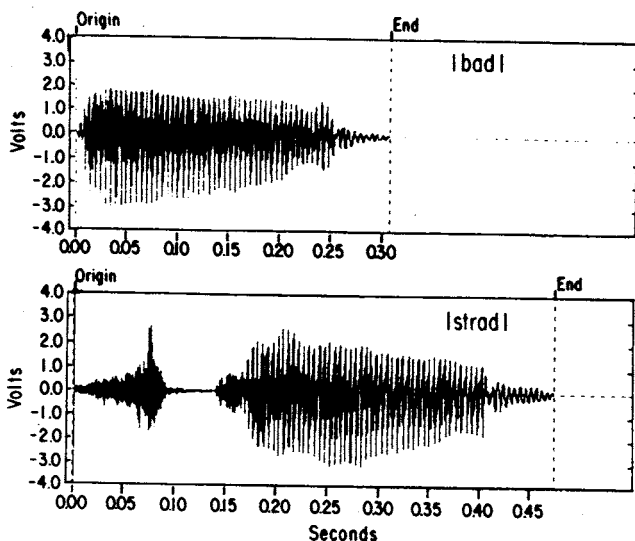


FIG. 1. Waveforms of the syllables used in experiments 1 and 2.

onset-onset time. Stress-beat-aligned sequences had a 465-ms interval between [b]-release and [s]-fricative, and a 735-ms interval between [s]-fricative and [b]-release. Clicks coincided with syllable onset (that is, stop release of [b] and fricative onset of [s]) or with the stress beats of the syllables (stop release of [b] and 135 ms into the [str] cluster). The clicks were 20 ms in duration with a rise time of 1.7 ms, and were produced electronically.

### 3. Procedure

Infants were seated in an infant car seat equipped with stereo speakers. The car seat was placed on a table in a sound-attenuating chamber; the seat faced a (motionless) toy bear. The infant's parent stood to the infant's right holding a bottle. Parents were blind with respect to the purposes of the study. The bottle's nipple was attached to a pressure transducer connected to an analog-to-digital converter interfaced to the computer. Sucks were detected by a pattern-recognition routine in the program that ran the experiment. The routine was designed after Kaye's criteria (1967) for identifying high-amplitude sucks.

The parents were instructed to keep the nipple in the infant's mouth unless the infant actively rejected it. They were instructed in the event of rejection to make the nipple continuously available to the infant. An experimenter (the second author) remained in the booth with the parent and infant to monitor the proceedings. The experimenter was not visible to the infant during the procedure. For half of the babies (approximately), parents wore headphones and listened to masking music during the session; for the remaining half, they did not. This manipulation was intended to reveal whether the parents' awareness of the stimulus conditions would affect the babies' sucking behavior. After a short period in which the parent and baby adjusted to the setting, a 1-min baseline interval began. In the interval, the child's sucking was monitored in the absence of any speech or clicks being presented over the speakers. Following the baseline minute, the speech sequences were presented at a low ampli-

tude over the speakers. {The peak vowel amplitude for both [bad] and [strad] was 75 dB SPL(A) at the speakers and about 15 dB less at the infant's ear level.} Clicks were presented over a separate speaker placed behind and slightly above the infant car seat, midway between the stereo speakers presenting the speech. Sucking was monitored and sucking rate was used to regulate the amplitude of the speech. Sucking was monitored in overlapping 3600-ms windows. An increase in sucking rate in one window as compared to the previous one caused a proportional increase in the amplitude of the speech up to a maximum [90 dB SPL(A) at the speakers]. Correspondingly, a decrease in sucking rate caused a proportional decrease in amplitude down to a minimum [75 dB SPL(A) at the speakers]. Clicks were invariant in amplitude and were 60 dB SPL(A) measured at the stereo speakers presenting the speech.

Beginning in minute five of the acquisition phase, the computer program checked for habituation. In minute five, the number of sucks in minute three was compared to those in the two subsequent minutes. If the number of sucks in both minutes four and five were 0.75 or less of the number of sucks in minute three (cf. Jusczyk, 1977; Jusczyk *et al.*, 1983), the infant was considered to have habituated and was shifted to the test phase of the study. If habituation had not occurred, the acquisition phase continued. Habituation was checked in an analogous fashion (comparing sucking in each minute  $x$ , with sucking in minutes  $x + 1$  and  $x + 2$ ) at the end of each minute following the fifth until habituation was detected or until the child had completed 16 acquisition-phase minutes. In either case, the child was shifted to the test phase.

The test phase lasted 3 min in which infants in the experimental conditions heard a new sequence and control infants heard the same sequence as in the acquisition phase. In this phase, as before, amplitude of the speech sequence was contingent on sucking rate.

For infants to contribute to our discrimination measure, they had to have shown habituation by the 16th minute and to have completed the session through the second test-phase minute. For them to contribute to the measure of time to habituate, they had to complete the session through the shift point between acquisition and test.

#### 4. Subjects

Ninety-one infants were run in the four experimental conditions and four control conditions of the experiment. Of these, data from 34 infants were excluded entirely from the analyses for a variety of reasons. Twenty-one babies were rejected from the experiment for state changes before the shift point (e.g., crying, falling asleep); six infants were rejected for not sucking at all or for failing to exceed their level of sucking during the baseline minute of the acquisition phase; and seven infants were lost from data analysis for reasons unrelated to the infants' own behavior (e.g., equipment failure).

Of the remaining 57 infants run in the experiment, data from 16 were excluded from the discrimination measure. Of these, 11 infants were run successfully through the shift point following habituation, but not for the 2 min thereafter

needed for the discrimination measure. The remaining five infants had not habituated by the end of the 16th acquisition-phase minute. Nine of the 11 infants who did not complete the two post-shift minutes were in the control condition. This imbalance is not surprising in view of the fact that control-condition infants did not receive a new speech sequence after habituation, whereas experimental-condition infants did. However, the differential dropout rate does mean that, despite random assignment of infants to conditions before the experimental manipulations were performed, after the shift point the groups may be different. This does not affect our time-to-habituate measure, but it could affect comparison of experimental- and control-condition infants on a discrimination measure. We will return to this point in Sec. I B.

Other than that asymmetry, there were no striking imbalances in numbers of infants lost from the eight conditions of the experiment. Of the 12 control-condition infants (nine who failed to complete the post-shift minutes and three who failed to habituate) lost from the discrimination measure, six were from the condition in which parents heard masking music; likewise, of the four experimental-condition infants lost (two who failed to complete the two post-shift minutes and two who failed to habituate), two were from each masking-music condition.

Infants ranged in age from 11 weeks, 0 days to 24 weeks, 4 days, averaging 17 weeks, 1 day. They were recruited by contacting parents whose names were listed in the birth notices in the local newspaper.

## B. Results

### 1. Performance profiles: Sucking rates with a contingency between speech amplitude and sucking rate

Figure 2 shows the performance profiles of those infants who remained in the experiment at least through the shift point. The profiles are plotted separately for infants receiv-

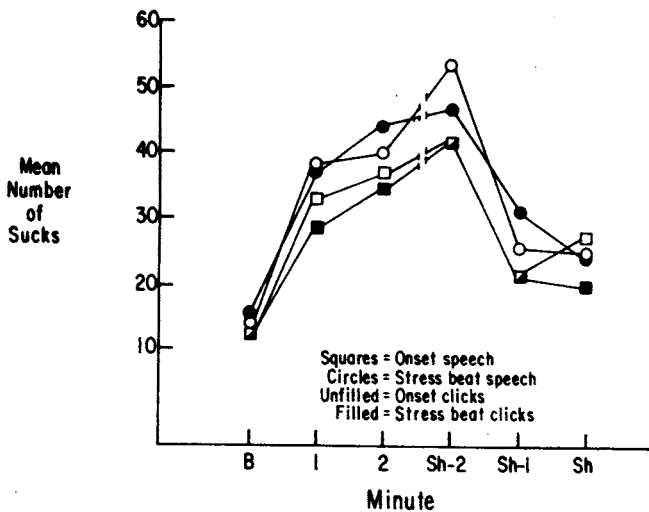


FIG. 2. Mean rates of sucking during the baseline (B), the first two acquisition-phase minutes, and the last three minutes before the shift (labeled "Sh-2," "Sh-1," and "Sh") for infants in the four acquisition-phase conditions of experiment 1. (See Table I for a description of the conditions.)

ing the four acquisition conditions displayed in Table I. Because this phase of the experiment preceded that in which experimental and control groups received different treatment, these groups were collapsed. This gave 14 infants in three of the acquisition-phase conditions and 15 in the fourth.

The figure displays the mean number of sucks per minute on the ordinate and six sampled minutes during the experiment on the abscissa: the baseline minute, the first two, and the last three acquisition-phase minutes. For infants who habituated at the fifth minute, this includes all of their acquisition-phase data. However, for infants habituating after the fifth minute, intermediate minutes after the second and before the last three are excluded from the figure.

An analysis of variance was performed on these data, with three between-group factors: speech timing (onset, stress-beat), relative click timing (onset, stress-beat), and masking music, and one within-group factor, minute (the 6 min displayed in the figure). The only factor to reach significance in the analysis was the main effect of minute [ $F(5,245) = 35.03, p < 0.001$ ]. Planned comparisons on the means contributing to this factor confirm that sucking rates increased significantly from the baseline during the first two acquisition-phase minutes [ $F(1,245) = 75.24, p < 0.001$ ]<sup>4</sup> and that, as a consequence of the habituation criterion, they fell during the last two, as compared to the preceding minute [ $F(1,245) = 87.93, p < 0.001$ ].

In this analysis, which focuses on the infants' sucking behavior during acquisition of the contingency and during habituation, infants hearing the four different acquisition-phase sequences do not behave significantly differently.

These performance profiles resemble those reported by other investigators using the more usual contingency between sucking rate and rate of stimulus presentation (cf. Jusczyk, 1981). Accordingly, we conclude that a contingency between speech amplitude and sucking rate can be used to study speech perception by infants in cases where speech timing is a manipulated variable.

## 2. Time to habituate

We turn now to the dependent measure of greatest interest in the experiment, number of minutes before habituation for infants hearing the four different speech-timing by click-timing sequences. The data were analyzed with the factors speech timing, relative click timing, and masking music. In each of the four speech-timing by click-timing cells, there were eight infants whose parents did not listen to masking music; of infants whose parents did listen to masking music, there were six in three of the cells and seven in the fourth. The mean times-to-habituate are presented in Fig. 3, collapsed over the factor of masking music. The main effects of masking music [ $F(1,49) < 1$ ] and of speech timing [ $F(1,49) = 2.19, p = 0.14$ ] were not significant. However, the effect of relative click timing was significant [ $F(1,49) = 5.66, p = 0.02$ ], reflecting longer acquisition phases for the two groups of infants hearing the sequences with clicks aligned with stress beats than for the two groups hearing the sequences with onset-aligned clicks.

There was one significant interaction in the analysis,

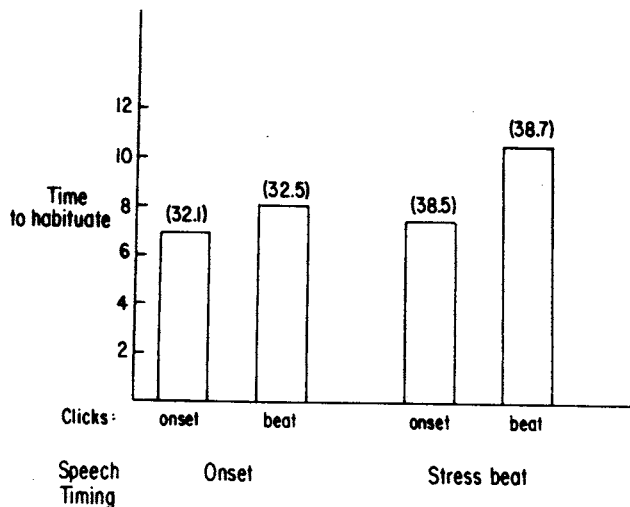


FIG. 3. Data from experiment 1. Mean times-to-habituate (in minutes) and, in parentheses, average rates of sucking before habituation for each of the four combinations of speech timing and relative click timing.

between masking music and speech timing [ $F(1,49) = 4.43, p = 0.038$ ]. Figure 3 reveals a tendency for stress-beat aligned speech to be associated with slower rates of habituation than onset-timed speech. This tendency was marked in the condition in which parents did not hear masking music, but was absent in the condition in which they did.

As a final analysis on the time-to-habituate measure, we performed a planned comparison to ask whether the condition that adults hear as isochronous with synchronous clicks was distinguished in the infants' sucking behavior from the other conditions. The analysis was performed on the eight means representing the crossing of masking music, speech timing, and relative click timing. Of these eight means, the two representing this condition differed significantly from the remaining six [ $F(1,49) = 7.71, p = 0.008$ ].

We also examined the average number of sucks per minute before habituation for infants hearing the four different acquisition-phase sequences. The measures are given in parentheses above each corresponding time-to-habituate bar in Fig. 3. The measures reveal an apparently higher sucking rate in conditions with stress-beat aligned speech than in conditions with onset-aligned speech. In addition, the conditions associated with the highest and lowest rates of sucking are the same on these measures as on the measure of minutes of sucking before habituation. However, differences in the sucking rates do not approach significance. The correlation between this measure and time-to-habituate is significant ( $r = 0.36, p = 0.003$ ), but small.

## 3. Discrimination

We obtained the discrimination measures in case infants failed to show differences in time-to-habituate across the experimental conditions. Had they nonetheless discriminated the sequences, one plausible explanation for their failure to show differential responsiveness to them, namely that they could not discriminate the sequences, would be ruled out. In the event, there were differences in time-to-habituate, and

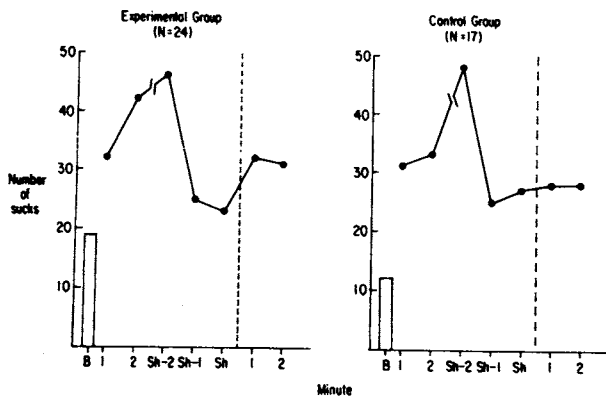


FIG. 4. Mean rates of sucking during the baseline (B) the first two acquisition phase minutes, the last three minutes before the shift point ("Sh-2," "Sh-1," and "Sh"), and the two minutes after the shift point. Data are plotted separately for infants in the experimental and control conditions of experiment 1.

this obviates dwelling on the discrimination measure. We present the data, however, for completeness and as a further indication that the amplitude contingency provides conventional performance profiles in a non-nutritive sucking, habituation paradigm.

As we noted in Sec. I A, relatively more control- than experimental-condition infants were lost from the discrimination measure. Although this is understandable in terms of the different treatments the two groups received at the shift point, it does create the possibility of group differences in performance unrelated to the change in timing of the speech sequence. Figure 4 helps to allay this concern somewhat. The figure plots performance profiles for those experimental- and control-condition infants who remained in the experiment through the second post-shift minute. The two groups look similar in the acquisition phase of the experiment—an observation confirmed by an analysis of variance performed on the preshift minutes displayed in the figure. The analysis had two between-group factors, group (experimental, control) and masking music, and one within-group factor, minute. As in the analogous analysis of the larger group of infants reported earlier, the only significant factor in the analysis was minute [ $F(5,185) = 28.81, p < 0.001$ ]. As in the earlier analysis, the significant effect of minute reflects the increase in sucking over the baseline within the first two acquisition-phase minutes and the decrease in sucking in the last two minutes of the acquisition-phase as compared to the preceding minute.

Our measure of discrimination compared sucking rate in the last minute before the shift point with the average of the last two minutes after the shift. Sequences presented to experimental-condition infants after the shift differed from preshift sequences in speech timing. The absolute timing of the clicks was not changed; therefore, the timing of the clicks relative to the speech shifted from onset- to stress-beat-aligned or vice versa. On this comparison, experimental-condition infants significantly increased their sucking rate from 23 sucks in the minute before the shift to 32 after the shift [ $t(23) = 3.40, p = 0.0025$ ]. Thus the change in speech timing after the shift and the consequent change in relative

timing of the clicks and speech was discriminated by experimental-condition infants. Control-condition infants increased their sucking rate nonsignificantly from 27 sucks per minute before the shift to 28 sucks per minute thereafter [ $t(16) = 0.32$ ]. In addition, we compared the increase of nine sucks per minute exhibited by experimental-condition infants with the one suck per minute increase by control infants in an analysis of variance with factors group and masking music. The only significant factor in the analysis was that of group [ $F(1,37) = 6.24, p = 0.016$ ].

### C. Discussion

Experiment 1 was designed to determine whether infants would show differences in the number of minutes they suck before habituating to sequences of speech syllables and clicks differing in their relative timing. The positive outcome we obtained establishes the conditions needed for a second experiment to interpret the differences in terms of the infants' perceptions of syllable timing and syllable stress beats.

As we have already pointed out, conclusions about the infants' perception of speech timing cannot be drawn with confidence from the results of experiment 1 alone. Possibly, infants hear syllable timing by reference to stress beats, and the longer times to habituate when clicks are on the stress beat may reflect the infants' greater interest in congruent over incongruent speech and click timing. Compatibly, their longest times-to-habituate associated with stress-beat-timed speech and clicks may reflect an interest in isochronous speech rhythms, perhaps because they are relatively natural or prototypical. However, another possibility, consistent with Kuhl's interpretation of her findings, is that long times to habituate are associated with more complex, varied stimuli. In that case, a likely inference is that infants do not hear speech timing by reference to stress beats. Therefore, the stress-beat-aligned clicks sound incongruent and the stress-beat-timed speech long-short; they lead to relatively long times-to-habituate because they are complex.

Experiment 2 is designed to circumvent these difficulties of interpretation by comparing the pattern of times-to-habituate in experiment 1 with another pattern of times-to-habituate generated by speech sequences whose perceived timing is unquestionably isochronous on the one hand and long-short on the other and with click patterns of unambiguously perceived relative timing. In this way, regardless of the reasons why the various speech-and-click sequences led to different times-to-habituate in experiment 1, their perceived timing can be discovered.

In addition, however, experiment 2 is designed to determine whether long times-to-habituate in experiment 1 are due to perceived complexity or to some other property of the stimulus sequence.

### II. EXPERIMENT 2

For the purposes of the second experiment, we created three new sequences of speech syllables with superimposed clicks. In the new sequences, only the syllable [strad] was presented. In two sequences, onset-onset times were isochronous; in the third, the syllables had a long-short alternating rhythm. Because only the syllable [strad] was used,



isochronous sequences are perceptually isochronous whether listeners attend to syllable onsets, to stress beats, or anywhere else in the syllables. Likewise, the long-short sequence has a perceptually long-short alternating rhythm whether onsets or stress beats are the focus of attention.

In one isochronous sequence and in the long-short sequence, clicks fell 135 ms within each [strad] syllable; that is, they fell on the stress beats. By comparing times-to-habituate in these two sequences, we can determine whether longer times-to-habituate are associated with the more complex, long-short rhythm or with the simpler, isochronous rhythm. If they are associated with the long-short rhythm, then Kuhl's interpretation of the variation in times-to-habituate she observed presumably applies to the present paradigm and findings as well. Infants habituate more slowly to sequences that are more difficult to encode. If, instead, the more uniform, isochronous rhythm is associated with longer times-to-habituate, then a more likely interpretation is that the isochronous rhythm was the more pleasing or interesting, perhaps because it corresponds more closely than the long-short rhythm to the underlying rhythm prototypical of spoken languages.

In the other isochronous sequence, clicks fell alternately on the onset of one [strad] syllable and 135 ms within the next. The set of three sequences—*isochronous speech with clicks on the stress beats, isochronous speech with clicks alternately on onsets and stress beats, and long-short speech with clicks on the stress beats*—will allow us to interpret the pattern of times-to-habituate observed in experiment 1. If infants attend to stress beats rather than to onsets of syllables in assessing syllable timing, then stress-beat timed syllables in experiment 1 will be perceived as isochronous, corresponding to the isochronous sequences of the present study, and onset-timed speech in experiment 1 will correspond to the long-short rhythm in experiment 2. In contrast, if infants attend to syllable onsets, then onset-timed speech in experiment 1 should correspond to isochronous speech in experiment 2 and stress-beat timed speech to the long-short sequences. The two click alignments in the isochronous sequences will test how the click timing of experiment 1 was perceived by infants. The different click placements in experiment 1 may be perceived in either of two ways, because the stress beat of [bad] is on the syllable's onset. One possibility is that infants in experiment 1 perceived syllable stress beats as adults do. In that case, the pair of sequences to which infants habituated slowly in experiment 1 had clicks perceived as consistently synchronized with the syllables, while the pair to which they habituated more rapidly were perceived as having clicks synchronized with [bad] but offset from the stress beat of [strad] (that is, having a "syncopated" rhythm). An alternative (among others) is that infants are insensitive to stress beats and hear syllable timing by reference to the timing of syllable onsets. In that case, the pair of sequences to which they habituated slowly in experiment 1 had a syncopated rhythm while the pair to which they habituated relatively rapidly had clicks perceived as uniformly synchronized with the syllables. The syncopated and uniform click-timing placements in experiment 2 were designed to test the first possibility (that infants are sensitive

to stress beats) against others, including the second.

We used a trend analysis designed to compare the patterns of times to habituate in experiments 1 and 2. The analysis first makes the prediction that infants hear syllable timing as adults do. If they do, the three conditions of experiment 2, isochronous speech with clicks on the stress beats, isochronous speech with clicks alternately on and off the stress beats, and long-short speech with clicks on the stress beats, should correspond most closely, respectively, to conditions in experiment 1 in which infants sucked 10.5 min (stress-beat aligned speech with clicks on the stress beats), 7.4 min (stress-beat aligned speech with clicks on the onsets), and 8.0 min (onset-aligned speech with clicks on the stress beats). To determine whether the pattern of sucking times in experiment 2 matches that for corresponding conditions in experiment 1, we perform a trend test on the three conditions of experiment 2 in which the sucking times are weighted according to the times we hypothesize correspond in experiment 1. If the weighting is appropriate, then the analysis should account for most of the systematic variability in sucking times in experiment 2. We can compare the analysis with means weighted in this way to analyses with other weightings that reflect other hypotheses about the infants' mode of attention to the speech signal.

## A. Method

### 1. Stimulus materials

As previously described, there were three acquisition-phase sequences. In all three, the syllable [strad] from experiment 1 was presented repeatedly. To accommodate the extra duration of [strad] as compared to [bad], the cycle duration of a syllable pair was increased to 1300 ms in the sequences. In two sequences, onset-onset times were 650 ms; in the remaining sequence, they were alternately 515 and 785 ms. In one isochronous sequence and in the long-short sequence, clicks fell 135 ms after the onset of [strad]; in the third sequence, they fell alternately on the onset of one [strad] and 135 ms after the onset of the next.

### 2. Procedure

The procedure was the same as described for experiment 1. All infants were shifted to a new sequence following habituation. However, we will not present discrimination data, because we have relatively few babies in the study and we did not run a no-shift condition. All parents in the experiment listened to masking music.

### 3. Subjects

Of the 22 infants run in the experiment, data from seven were eliminated from analysis based on the same criteria used in experiment 1. This left five infants in each condition. Of infants rejected from the experiment, five were rejected for crying and two for failing to suck. Infants ranged in age from 14 weeks, 4 days to 21 weeks, 5 days, averaging 17 weeks, 6 days.



The first analysis examined sucking rates during the acquisition phase. Factors were group and minute (baseline, the first two, and the last three acquisition-phase minutes). As in experiment 1, the only factor to reach significance was minute [ $F(5,60) = 13.07, p < 0.001$ ]. Its significance was due, as earlier, to the increase in sucking in the first two acquisition-phase minutes over the baseline [ $F(1,60) = 34.03, p < 0.001$ ] and to the decrease in sucking during the last two minutes as compared to the preceding acquisition-phase minute [ $F(1,60) = 29.82, p < 0.001$ ].

On the measure of time-to-habituate, infants sucked 10.0 min, on the average, to isochronous speech with clicks on the stress beats. They sucked 6.4 min to isochronous speech with clicks alternating between onset and stress beat, and 6.8 min to short-long speech with clicks on the stress beats.

One notable outcome is that the short-long sequence in experiment 2 was associated with a shorter time-to-habituate (6.8 min) than the comparable isochronous sequence (10 min). By most accounts, a short-long rhythm is a more complex, varied rhythm than an isochronous rhythm; therefore, the present findings offer no support for the idea that long times-to-habituate in our paradigm reflect encoding difficulty as they appeared to in the study by Kuhl (1976).

Turning to a comparison of times-to-habituate among the three conditions of experiment 2, we note that they are ranked identically to those in experiment 1 under our hypothesis that stress-beat-aligned speech is isochronous to infants and that clicks are perceived as synchronized to speech when they fall on the stress beat. Indeed, the differences among the three conditions are significant in a planned comparison in which the means are weighted according to the different conditions that we hypothesize are analogous in experiment 1 [ $F(1,12) = 6.44, p = 0.025$ ].<sup>5</sup> This analysis accounts for essentially all of the systematic variability in the data. That is, in the overall analysis of variance, the main effect of condition accounted for 23% of the total variability in the data, a strong effect. Our theoretically motivated weighting of the means captures more than 99% of that systematic variability in the main effect.

Obviously, no other hypothesis predicting different relations among the conditions of experiments 1 and 2 could do as well. For example, a hypothesis that onset-aligned speech sounds isochronous to infants would pair the 8-min condition in experiment 1, in which speech was onset aligned with clicks on the stress beats, with the 10-min condition in experiment 2, and the 10.5-min condition in experiment 1, in which onsets were long-short (because stress beats were isochronous) with the 6.8-min condition in experiment 2. That is, it would pair the longest sucking time in experiment 1 with the next longest in experiment 2 and vice versa.

Based on the results of the planned comparison, we infer that the sequence in experiment 1 associated with the relatively slow rates of habituation sounded isochronous to infants and that the timing of clicks relative to speech associated with relatively slow rates of habituation in experiment 1 is on the stress beats, rather than alternately on and off syllable onsets.

There are four major findings of our experiments. First, in experiment 1, infants showed differential rates of habituation to onset- and stress-beat timed syllable sequences on which clicks were superimposed either on measured syllable onset or on the syllables' stress beats. Second, when infants' rates of habituation are compared among sequences whose perceived relative timing is known (in experiment 2), infants habituated more slowly to isochronous sequences with clicks on the stress beats than to sequences with different speech timing or to sequences with different click placements. Third, and most interesting, this pattern of differential rates of habituation in experiment 2 was accounted for essentially completely by assumptions that stress-beat-aligned syllables in experiment 1 were isochronous to infants, while onset-aligned syllables had a long-short rhythm, and that clicks on the stress beats in experiment 1 were perceived as synchronized to the speech whereas clicks on the onsets were not. In turn, this implies that infants hear speech timing by reference to timing of stress beats just as adults do. A fourth finding is more peripheral to the main purpose of the study. It is that times-to-habituate in the present paradigm do not appear to reflect encoding difficulty of the stimulus sequences. Most likely, infants habituate slowly to sequences that interest them. Their interest may be provoked by novelty, encoding difficulty or, as here, perhaps relative pleasingness of the stimulus sequences.

In reference to the main findings of the study, our interpretation of the corresponding findings on adults (Fowler, 1979; Tuller and Fowler, 1980; Fowler and Tassinari, 1981) is that adults attend to information about articulatory timing. Although, as we indicated in the Introduction, this interpretation requires further testing and there are competing views, it is the only approach to date to uncover anything isochronously timed in perceptually isochronous productions, and the only one to rationalize the stress beat. If the interpretation is correct, it implies that infants, like adults, use the acoustic speech signal as information about its articulatory source.<sup>6</sup>

In a sense, this should not be a surprising conclusion. In order to learn to speak from adult models, infants must extract gestural information from the acoustic speech signal. Therefore, eventually, they must treat the acoustic signal as a carrier of information about its articulatory source. On this interpretation, our research suggests that infants do that even before they have experience during babbling of hearing the acoustic consequences of their own speechlike articulations.

#### ACKNOWLEDGMENTS

This research was supported by NICHD grant HD 16591-01 to Haskins Laboratories. We thank Alvin Liberman, D. K. Oller, Michael Studdert-Kennedy, and George Wolford for their comments on earlier drafts of the manuscript.

## APPENDIX

In our first pilot study, we determined the timing of pairs of syllables that sounded isochronous to adults. We tested three pairs of monosyllables that previous research (Fowler and Tassinary, 1981) had indicated would show large departures from measured acoustic isochrony of acoustically defined onsets when listeners achieved perceptual isochrony. The pairs were naturally produced [bad sad], [bad šad], and [bad strad]. In the pilot study, we used a procedure that we call the "method of adjustment" (cf. Cooper *et al.*, 1984). In the procedure (cf. Morton *et al.*, 1976; Marcus, 1981), listeners hear two syllables of a pair presented alternately in repeated cycles and they adjust the relative timing of the syllables within a fixed time window. Subjects are instructed to adjust the timing of the syllables until the sequence sounds isochronous (that is, as if it were timed to the beat of a metronome).

In the pilot study, each syllable-pair cycle occupied 1200 ms. The pairs were presented under computer control with an initial interval of 50 ms between the offset of acoustic energy of one syllable and the onset of acoustic energy of the next. Two listeners (the first two authors) adjusted the timing of each syllable pair by pressing keys on the calculator pad of a computer terminal keyboard. Their button pressing moved the second syllable in time relative to the first. They adjusted each pair four times in each of the two possible orders of presentation of the syllables, giving eight adjustments per syllable pair per subject. The two subjects showed very close agreement in the onset-onset asynchronies yielding perceived isochrony across the three syllable pairs. The results are shown in Fig. A1 presented as differences between A-B and B-A intervals, where A and B are the two syllables in the cycle. We refer to this difference as the "onset-onset" difference. The syllable pair [bad strad] showed the largest onset-onset differences among the three tested; therefore, we favored it as the pair to present to the infants.

Our second pilot study had two purposes. First, it tested whether naive listeners would hear clicks coincident with syllable onsets as unsynchronized with the speech and hear clicks on the stress beats as synchronized. Stress-beat alignment was determined from results of the first pilot study and findings by Fowler and Tassinary (1981). The first pilot study provided quantitative differences in the relative timing of sequences including [sad], [šad], and [strad], but it did not pinpoint where in the syllables a click should fall absolutely to sound synchronized with them. Fowler and Tassinary had found that [bad] spoken repeatedly in time with a metronome pulse is produced so that [b]-release coincides with the pulse. We used this placement of the pulse on [bad], along with the onset-onset anisochronies from the first pilot study, to obtain tentative locations for the clicks in [sad], [šad], and [strad]. Thus the click was located 109 ms into [sad], 116 ms into [šad], and 135 ms into [strad]. These placements conform closely with locations of the metronome pulse on [sad] and [strad] syllables in Fowler and Tassinary (1981).

A second purpose of the study was to select among alternative ways of presenting the click stimuli. We compared three manners of presentation. In two, clicks were presented

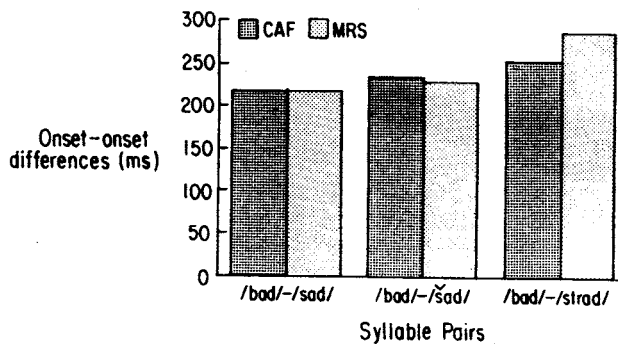


FIG. A1. Results of the first pilot study on adults. Mean onset-onset differences in perceptually isochronous sequences as a function of syllable pair for two adult listeners (CAF, MRS).

by a stuffed bear banging a drum under computer control. In one of those conditions, the bear's movements were visible to subjects; in another, they were not. In a third condition, clicks were presented by an electronic click generator under computer control. The click generator produced a sharper, shorter signal than the bear's drum beats.

Subjects in this study were 12 Dartmouth College undergraduates, naive to the purposes of the experiment. They heard a pair of sequences on each trial. The sequences in a trial consisted of the same pair of syllables (one of the three used in the first pilot study) produced several times in alternation, with one of two timing rhythms. The sequences were either onset-aligned or they had the onset-onset anisochrony found in the first study to yield perceptual isochrony. The sequences in a pair differed in the timing of the drum beats or clicks relative to the speech. In one sequence in each trial, clicks coincided with syllable onsets; in the other, they coincided with stress beats. All variables were manipulated within subjects. Subjects provided two judgments. First, they judged which of the two sequences had synchronous clicks and speech. Second, they provided a confidence judgment (1: confident; 2: somewhat sure; 3: random guess).

The results were clear. Subjects were more consistent (and more confident) in conditions involving computer-generated clicks rather than those involving the drum beats. They were least consistent when clicks were provided by the bear when it was visible. In addition, subjects were somewhat more consistent on [bad strad] sequences than on others, although differences were not large. In the best condition ([bad strad] with computer-generated clicks), subjects averaged nearly 90% agreement in choosing clicks aligned with stress beats as synchronous. Their confidence averaged 1.17.

Accordingly, we selected the [bad strad] sequence for use with the infants. We used the onset-onset difference (rounded to 270 ms) found in the first pilot study to yield perceived isochrony and the click timing found in the second pilot study to sound synchronous to adults. In other conditions, we used [bad strad] sequences with isochronous intervals between syllable onsets, measured from the acoustic signal, and click presentations synchronized with syllable onsets.

<sup>1</sup>In this paragraph and elsewhere, reference to "measured" durations, unless otherwise indicated, signifies measurement based on a waveform or spectrographic display. Details of criteria for measurement can be found in the cited articles. Generally, measurements of prevocalic consonants begin at the onsets of acoustic energy for the syllable and end at measured vowel onset. Vowel onset begins at the onset of regular pulsing for the vowel following evidence of release of the consonantal constriction and ends at evidence of closure for the following consonant.

<sup>2</sup>Although Howell (1984) claims to have identified an acoustic correlate in the amplitude envelope of a word or syllable, research both previous to his (Marcus, 1981; Tuller and Fowler, 1981) and subsequent to it (Cooper *et al.*, 1984) disconfirms his claims.

<sup>3</sup>In the syllable [bad], voice-onset time (VOT) was 8 ms in duration, voicing before closure 244 ms, and voiced closure 56 ms. In the syllable [strad], friction for [s] was 96 ms in duration, closure for [t] 46 ms, VOT 29 ms, voicing prior to closure 237 ms, and voiced closure 66 ms. Neither syllable had a release burst. These measurements conform well to those of stimuli used with adult listeners (Fowler, 1979) and to those of stimuli produced by talkers under isochrony instructions (Fowler and Tassinary, 1981). As the pilot studies in the Appendix reveal, they lead to conventional judgments among adult listeners.

<sup>4</sup>Without removing the contingency, of course, we cannot conclude from these findings that the increase in sucking reflects acquisition of the contingency rather than arousal or some other factor. This same ambiguity is present in most uses of this type of paradigm. However, it is likely in the present study that infants did learn the contingency. Research by Williams and Golenski (1978) shows that removal of the contingency in the paradigm of Eimas *et al.* (1971) eliminates differences between shift and no-shift groups and increases dropout rates. A recent unpublished study of our own replicates the latter observation, when the paradigm of the present experiments is used, and finds no differences in time-to-habituate between groups hearing the most and least favored sequences of the present study. Accordingly, we are inclined to interpret the sucking profiles in experiment 1 as an indication that infants did learn the contingency.

<sup>5</sup>We used a procedure suggested to us by George Wolford (personal communication) and described in Wolford and Fowler (1983). It is a generalization of the procedure for detecting linear or quadratic trends in data (cf. Hayes, 1973). To detect a linear trend in a set of three means, for example, the means are given the weights -1, 0, and 1 in a planned comparison. However, the weights may be selected to reflect any theoretically motivated spacing of the means. In our analysis, the weighting was determined as follows. The grand mean of the relevant sucking times, 10.5, 8.0, and 7.4 min, from experiment 1, was calculated. The deviation of each of these sucking-time means from their grand mean yielded the weights 1.87, -0.63, and -1.24, respectively. These weights reflect the pattern of separations among the three relevant conditions of experiment 1 and, as required by the trend analysis, they sum to zero. Applied as weights to the means of experiment 2, in accordance with the hypothesis that infants hear syllable timing as adults do, 1.87 is the weight for the condition in which infants suck to hear isochronous speech with clicks on the stress beats, -0.63 is the weight for the condition in which they suck to hear long-short speech with clicks on the stress beats, and -1.24 is the weight for the remaining condition (isochronous speech with clicks on and off of the stress beat).

<sup>6</sup>We wish to forestall the inference made initially by our reviewers that two consequences of our interpretation are that speech perception by reference to articulation is "prewired" and that it is necessarily, thereby, different from auditory perception in general. We are agnostic on both points. We are testing three- to six-month-old infants, and therefore have no information concerning prewiring. Our claim is simply that infants do not need to have heard their own canonical babbling in order to hear syllable timing as adults do. In our view (Fowler, 1984, in press), a claim that perceivers use the acoustic speech signal as information about its distal articulatory sources makes speech perception look very much like perception in any modality, including general auditory perception and visual perception. Perception, in general, is a process of extracting information in "proximal stimulation" available at the sense organ about distal, environmental events. Accordingly, we might expect the P center itself not to be special to speech. Rather, it may be found to characterize any audible events in which the onset of acoustic energy created by a sound-producing event does not coincide temporally with the onset of the event itself in the environment.

Allen, G. (1972a). "The location of rhythmic stress beats in English, I," *Lang. Speech* 15, 72-100.

Allen, G. (1972b). "The location of rhythmic stress beats in English, II," *Lang. Speech* 15, 179-195.

Allen, T., Walker, K., Symonds, L., and Marcell, M. (1977). "Intrasensory and intersensory perception of temporal sequences," *Dev. Psychol.* 13, 225-229.

Cooper, A., Whalen, D., and Fowler, C. (1984). "Stress centers are not perceived categorically," Paper presented at the 108th Meeting of the Acoustical Society of America, *J. Acoust. Soc. Am. Suppl.* 1 76, S89.

Demaney, L., MacKenzie, B., and Vurpillot, E. (1977). "Rhythmic perception in early infancy," *Nature* 266, 718-719.

Dodd, B. (1979). "Lip reading in infants: Attention to speech presented in- and out-of synchrony," *Cog. Psychol.* 11, 478-484.

Eimas, P., Siqueland, E., Jusczyk, P., and Vigorito, J. (1971). "Speech perception in infants," *Science* 171, 303-306.

Eling, P., Marshall, J., and van Galen, G. (1980). "Perceptual centres for Dutch digits," *Acta Psychol.* 46, 95-102.

Fowler, C. (1979). "Perceptual centers in speech production and perception," *Percept. Psychophys.* 25, 375-388.

Fowler, C. (1983). "Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet," *J. Exp. Psychol.* 112, 386-412.

Fowler, C. (1984). "Segmentation of coarticulated speech in perception," *Percept. Psychophys.* 36, 359-368.

Fowler, C. (in press). "An even approach to a theory of speech perception from a direct-realist perspective," *J. Phonetics*.

Fowler, C., and Tassinary, L. (1981). "Natural measurement criteria for speech: The anisochrony illusion," *Attention and Performance, IX*, edited by J. Long and A. Baddeley (Erlbaum, Hillsdale, NJ).

Hayes, B. (1980). "A metrical theory of stress rules," Doctoral dissertation, Massachusetts Institute of Technology.

Hayes, W. (1973). *Statistics for the Social Sciences* (Holt, Rinehart and Winston, New York), 2nd ed.

Hoequist, C. (1983). "The perceptual center and rhythm categories," *Lang. Speech* 26, 367-376.

Howell, P. (1984). "An acoustic determinant of perceived and produced anisochrony," *Proceedings of the Tenth International Congress of Phonetic Sciences*, edited by M. P. R. Van den Broecke and A. Cohen (Foris, Dordrecht, Holland).

Humphrey, K., Tees, R., and Werker, J. (1979). "Auditory-visual integration of temporal relations in infants," *Can. J. Psychol.* 33, 347-352.

Jusczyk, P. (1977). "Perception of syllable-final stop consonants by two-month-old infants," *Percept. Psychophys.* 21, 450-454.

Jusczyk, P. (1981). "Infant speech perception: A critical appraisal," in *Perspectives on the Study of Speech*, edited by P. H. Eimas and J. Miller (Erlbaum, Hillsdale, NJ).

Jusczyk, P., Pisoni, D., Reed, M., Fernald, A., and Myers, M. (1983). "Infants' discrimination of the duration of a rapid spectrum change in non-speech signals," *Science* 222, 175-177.

Kaye, H. (1967). "Infant sucking behavior and its modification," in *Advances in Child Development and Behavior*, edited by L. Lipsitt and C. Spiker (Academic, New York), Vol. 3.

Kuhl, P. (1976). "Speech perception in early infancy: The acquisition of speech-sound categories," in *Hearing and Davis: Essays in Honor of Hallowell Davis*, edited by S. Hirsh, D. Eldredge, I. Hirsh, and S. Silverman (Washington U. P., St. Louis, MO.).

Kuhl, P., and Meltzoff, A. (1982). "The bimodal perception of speech in infancy," *Science* 218, 1138-1144.

MacKain, K., Studdert-Kennedy, M., Spiker, S., and Stern, D. (1983). "Infant intermodal speech perception is a left hemisphere function," *Science* 219, 1347-1349.

Marcus, S. (1981). "Acoustic determinants of perceptual centers (P-centers)," *Percept. Psychophys.* 30, 247-256.

Morton, J., Marcus, S., and Frankish, C. (1976). "Perceptual centers," *Psychol. Rev.* 83, 405-408.

Oller, D. K. (1980). "The emergence of the sounds of speech in infancy," in *Child Phonology: Production*, edited by G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (Academic, New York).

Prince, A. (1983). "Relating to the grid," *Linguistic Inquiry* 14, 19-100.

Rapp, K. (1971). "A study of syllable timing," *Pap. Linguistics U. Stockholm* 8, 14-19.

Spelke, E. (1976). "Infants' intermodal perception of events," *Cog. Psychol.* 8, 553-560.

Spelke, E. (1979). "Perceiving bimodally specified events in infancy," *Dev. Psychol.* 15, 626-636.

**Tuller, B., and Fowler, C. (1980).** "Some articulatory correlates of perceptual isochrony," *Percept. Psychophys.* 27, 277-283.

**Tuller, B., and Fowler, C. (1981).** "The contribution of amplitude to the perception of isochrony," *Haskins Laboratories Status Reports on Speech Research SR-65*, 245-250.

**Williams, L., and Golenski, J. (1978).** "Infant speech sound discrimina-

tion: The effects of contingent versus noncontingent stimulus presentation," *Child Dev.* 49, 213-217.

**Wolford, G., and Fowler, C. (1983).** "The perception and use of information by good and poor readers," in *Perception, Cognition, and Development: Interactional Analyses*, edited by T. Tighe and B. Shepp (Erlbaum, Hillsdale, NJ).