Perception & Psychophysics
2000, 62 (2), 333-340

Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns

LAUREL J. TRAINOR and BETH ADAMS
McMaster University, Hamilton, Ontario, Canada

Adults and 8-month-olds were presented with sequences in which every third complex tone was either longer or more intense. Segmentation was measured by comparing the detection of silent gaps inserted into three possible locations in each pattern: Silent gaps inserted at perceived segmentation boundaries are harder to detect than gaps within perceived phrases or groups. A go/no-go conditioned head-turn (hand-raising for adults) procedure was used. In Experiment 1, detection was worse for the gaps following the longer complex tones than for the gaps at the other locations, suggesting that the longer tones marked the ends of perceived groups for both infants and adults. Experiment 2 showed that an increase in intensity did not result in any systematic grouping at either age.

The perception of speech and music involves analyzing the structure of a pattern or stream of sounds that unfolds in time. One of the basic steps in this process is the segmentation of the stream into meaningful units or groups, such as notes, words, and phrases. For adults, it is likely that the segmentation of both music and language involves the use of both structural and acoustic cues. Structural cues in language include grammatical rules and phonotactic constraints. For example, some phonemes or groups of phonemes are more likely to begin words, whereas others are more likely to end words. Structural cues in music include the functional relations between the different notes of the scale. For example, musical compositions most often end on the tonic, or first note, of the scale. The acoustic cues to phrase endings in both speech (see, e.g., W. E. Cooper & Paccia-Cooper, 1980; W. E. Cooper & Sorensen, 1977; Klatt, 1976; Martin, 1970; Scott, 1982; Streeter, 1978) and music (see, e.g., Boltz, 1993; Clarke, 1985; Deutsch, 1980; Jusczyk & Krumhansl, 1993; Kidd, Boltz, & Jones, 1984; Todd, 1985) include a lengthening of the final element(s) of the phrase, a decrease in intensity at the end of the phrase, and changes in pitch. In performance, musicians use duration and intensity to make the segmentation clear (Drake & Palmer, 1993).

Adults appear to be predisposed to segment all auditory patterns; they will typically even impose a rhythm on isochronous sequences of identical elements, most often segmenting the elements into groups of two or three (Bolton, 1894; Woodrow, 1909). Studies of repeating tone patterns show that both duration and intensity affect adults' perceived segmentation (Fraisse, 1982; Handel, 1989; Povel & Essens, 1985; Povel & Okkerman, 1981; Preusser, Garner, & Gottward, 1970; Royer & Garner, 1966). It has been proposed that increasing the duration of every second, third, or fourth element in an otherwise isochronous series of identical elements results in a segmentation, with the longer elements marking the ends of groups (Fraisse, 1956, 1982; Woodrow, 1909, 1951). On the other hand, a number of researchers have suggested that increasing the intensity of every second, third, or fourth element most often results in a segmentation in which the more intense elements mark the beginnings of groups (Fraisse, 1982; Handel, 1989; Woodrow, 1951). However, the evidence that adults place segmentation boundaries after longer elements and before more intense elements in a sequence of otherwise identical isochronous elements is often anecdotal, based on self-report, or flawed methodologically (by always presenting sequences that start at the beginning of the phrase listeners were expected to hear). Thus, it is important to replicate these findings.

Some time in the 2nd half-year of life, infants appear to master the segmentation problem in both speech (see, e.g., Echols, Crowhurst, & Childers, 1997; Hirsh-Pasek et al., 1987; Jusczyk, Cutler, & Redanz, 1993; Jusczyk et al., 1992; Morgan, 1994, 1996; Morgan & Demuth, 1996) and music (Jusczyk & Krumhansl, 1993; Krumhansl & Jusczyk, 1990). Infants may initially have only acoustic cues to segmentation at their disposal, since they lack knowledge of grammatical structure in speech and of scale structure in music (Trainor & Trehub, 1992; Trehub & Trainor, 1993; Trehub, Trainor, & Unyk, 1993). Indeed, a number of researchers have proposed that infants use acoustic cues as a bootstrap into language seg-

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Science and Engineering Research Board of McMaster University to L.J.T. We are grateful to Jacqui Atkin and Adrienne Rock for testing the subjects and to Renée Desjardins and Daphne Maurer for commenting on an earlier draft. Correspondence concerning this article should be addressed to L. J. Trainor, Department of Psychology, McMaster University, Hamilton, ON, L8S 4K1 Canada (e-mail: ljt@mcmaster.ca).

—Accepted by previous editor, Myron L. Braunstein

mentation (e.g., Gleitman, Gleitman, Landau, & Wanner, 1988; Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Morgan, 1986; but see, also, Gerken, Jusczyk, & Mandel, 1994; Saffran, Aslin, & Newport, 1996). The question remains, however, as to how well infants are able to take advantage of these cues in the absence of other structural information. Thus, we chose to examine segmentation in simple tone patterns with neither musical nor linguistic structure.

Young infants are sensitive to rhythm and prosody, both of which are based, to some extent, on duration and intensity variations. With musical stimuli, infants are able to discriminate by 2 months of age (Demany, McKenzie, & Vurpillot, 1977) and to categorize by 7 months (Trehub & Thorpe, 1989) simple rhythmic patterns. With speech, infants can discriminate utterances differing only in their rhythm or stress patterns (Jusczyk & Thompson, 1978; Spring & Dale, 1977) and can perceive stress beats as adults do (Fowler, Smith, & Tassinary, 1985) within the first few months of life. In consonant–vowel strings, infants can perceive vowel duration changes (Eilers, Bull, Oller, & Lewis, 1984), as well as vowel intensity differences as small as 2 dB (Bull, Eilers, & Oller, 1984) in the 2nd half-year of life. Infants as young as 4 days of age can discriminate their native from a foreign language, presumably on the basis of prosodic features (Mehler et al., 1988). Furthermore, across the 1st year of life, infants prefer to listen to the exaggerated prosody of infant-directed over adult-directed speech (see, e.g., R. P. Cooper & Aslin, 1990; Fernald & Kuhl, 1987) and singing (Trainor, 1996), and caregivers tend to exaggerate the prosodic cues to segmentation when addressing infants (e.g., Bernstein Ratner, 1986; Fernald & Mazzie, 1991; Fernald & Simon, 1984; Morgan, 1986; Trainor, Clark, Huntley, & Adams, 1997).

In sum, infants' sensitivity to prosody and rhythm in linguistic and musical stimuli suggests that mechanisms for using duration and intensity to segment simple patterns may be in place during the infancy period. This study examines whether 8-month-old infants use an increase in duration to mark the ends of groups and an increase in intensity to mark the beginnings of groups in simple patterns composed of complex tones. This age group was chosen for the initial study because previous data suggest that they have already mastered many of the segmentation problems in speech and music. Once the role of duration and intensity is understood at this time point, younger infants could be examined.

A gap detection methodology was used in order to objectify the measurement of listeners' segmentation into phrases or groups and to allow the testing of nonverbal infants. This methodology has been used previously with infants. Specifically, Thorpe and Trehub (1992) presented infants with repeating six-tone sequences, where the first three elements differed from the last three elements in either timbre or pitch. The infants showed evidence of the so-called *duration illusion*, in that they found silent gaps inserted between same-pitch or same-timbre tones easier to detect than silent gaps inserted between different-pitch or different-timbre tones. These results provide evidence that infants segmented the six-tone patterns into two groups of three tones, where the tones in each group were either the same in pitch or the same in timbre.

In the present study, sequences were created in which complex tones were identical and isochronous, with the exception that every third complex tone was either longer (Experiment 1) or more intense (Experiment 2). The task was to detect the insertion of a silent gap at each of the three possible between-tone locations (e.g., in Experiment 1, between two short tones, a short followed by a long tone, or a long tone followed by a short tone). The rationale behind the gap detection methodology is that temporal information is better retained within than between perceived groups or phrases. Thus, it is possible to infer where listeners are placing segmentation boundaries, because gap detection performance at perceived boundaries would be expected to be poor. On the other hand, the detection of gaps not located at perceived segmentation boundaries would be expected to be relatively good. For example, if listeners placed segmentation boundaries after the longer element in a sequence where every third element was longer, gap detection performance would be expected to be lower for the gap located after the long element than at the other two possible locations.

EXPERIMENT 1

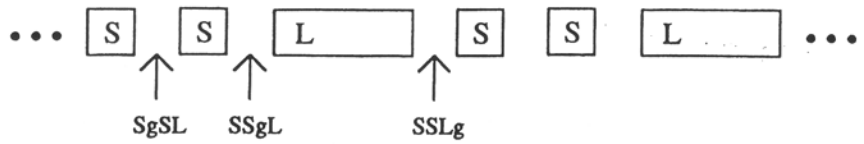
Method

Participants. Thirty 8-month-old infants (mean age = 8 months 13 days; 14 male, 16 female) were tested (10 per condition). All the infants were born within 2 weeks of term, weighed at least 2,500 g at birth with no known abnormalities, and all were healthy at the time of testing. One further infant was eliminated for failing to meet the training criterion. Thirty adults (mean age = 22 years; 9 male, 21 female) with no known hearing loss were also tested. None was a professional or a serious amateur musician (22 had between 0 and 5 years of music lessons; 8 had over 5 years).

Stimuli. A 256-Hz complex tone repeated continuously throughout the test session (see Figure 1a). The complex tone was composed of the first 10 harmonics, with a fall off in intensity of 6 dB per octave, and was presented at approximately 60 dB(A). Every third tone was 600 msec in duration, and the other two were 200 msec in duration, including 10-msec rise and decay times. All the tones were separated by 200 msec, so the onset-to-onset, in milliseconds, of successive tones followed the pattern ... 400, 400, 800, 400, 400, 800 ... (i.e., ... short–short–long–short–short–long ...). The 400-msec onset-to-onset duration falls within the range of syllable duration and is also close to the proposed "ideal" onset-to-onset duration of 500 msec (Handel, 1989, p. 386), at which adults show the most accurate duration discrimination. The onset-to-onset duration was doubled for the longer tones, since there is evidence that people tend to generalize a variety of beat lengths into two categories, where one is double the other (Fraisse, 1956; Povel, 1981).

There were three conditions, each starting on a different tone of the pattern, which will be referred to as SSL (short–short–long–short–short–long ...), SLS (short–long–short–short–long–short ...), and LSS (long–short–short–long–short–short ...). Note, however, that once the patterns started, the stimuli in the three condi-

a. Duration (Experiment 1)



b. Intensity (Experiment 2)

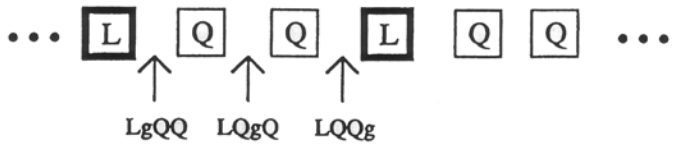


Figure 1. The stimuli and locations of the added silent gaps (g) to be detected in Experiments 1 and 2. In panel a, L refers to long, S to short; in panel b, L refers to loud, Q to quiet.

tions were indistinguishable. Ten infants and 10 adults were tested on each of the three starting tone conditions.

The task was to detect an increase in one of the silent gaps between complex tones (see Figure 1a). There were four types of trials—a control trial involving no added gap and three types of change trials: SgSL (gap inserted between short and short), SSgL (gap inserted between short and long), and SSLg (gap inserted between long and short). It was expected that the subjects would group the sequence as [short–short–long], regardless of the initial starting note (SSL, SLS, or LSS), and that performance would, therefore, be worse on SSLg than on either SgSL or SSgL. The duration of the tones that mark the beginning and the end of a silent period affect the threshold for detecting changes in the length of that silent period for short tone durations. Thus, in order to interpret the data in terms of grouping, it is important to choose tone durations that are expected to have no effect on gap discrimination. The 200-msec tone and the silent intervals were chosen to be within the range where little effect of the marker tones on gap detection was expected (Abel, 1972).

After pilot testing to determine the levels of difficulty for infants and adults that would result in above-chance, but not ceiling, performance, the gap increase was set to 100 msec for infants and 30 msec for adults.[1] During training, the gap increase was 200 msec for both the infants and the adults.

Apparatus. Testing was conducted in a sound-attenuating chamber (Industrial Acoustics Co.). The experiment was run using a Macintosh IIci computer, with an Audiomedia card to present 16-bit sounds and a Strawberry Tree I/O card to connect to a custom-built interface box. The interface box was connected in turn to a button box, lights, and mechanical toys located in the booth. The sound stimuli were passed through a Dennon amplifier (PMA-480) to a single loudspeaker (GSI) located inside the booth. The adult subject or parent holding the infant sat in a chair arranged so that the loudspeaker was located on the subject's left. Under the loudspeaker was a box with a smoked Plexiglas front containing four compartments, each in turn containing lights and a mechanical toy. It was not possible to see into the box except during reinforcement (see the Procedure section), when the lights and the toy in one compartment were turned on. The experimenter was seated across from the subject, behind a small table that concealed the button box.

Procedure. The infants were tested individually with a go/no-go conditioned head-turn response procedure. In each of the three conditions, the infants sat on their parent's lap in the sound-attenuating booth across from the experimenter. The experimenter and the parent listened to masking music through headphones and so were unaware of what the infant was hearing. The sequence repeated continuously throughout the test session. When the infant was attentive and facing straight ahead (i.e., toward the experimenter), the experimenter pressed a button to indicate to the computer that the infant was ready for a trial. Each infant received 40 trials: 10 were control trials; 30 were change trials, with 10 each of SgSL, SSgL, and SSLg. On control trials, the sequence continued to repeat without change. On each of the three types of change trials (SgSL, SSgL, and SSLg), a single gap increase was introduced (see the Stimuli section).

The infants were trained (see below) to turn their head toward the loudspeaker when a gap was introduced. The experimenter indicated to the computer via a second button whenever the infant turned at least 45° to the left to face the loudspeaker sound source. Thus, the experimenter's task was to call for a trial when the infant was attentive and centered and to indicate whenever the infant turned toward the speaker. The computer program automatically ran the rest of the procedure. Turns that occurred on change trials within 2,000 msec from onset of the inserted gap were reinforced by turning on one of the animated toys and the associated lights in the toy box under the loudspeaker for 1,600 msec. Turns at other times were not reinforced. The computer also recorded any head turns occurring within the same 2,000-msec time window on control trials (false alarms), to provide a measure of the rate of random turning. The trials were presented in a different quasi-random order for each subject, with the constraint that no more than two control trials occurred sequentially. As well, a minimum of six repetitions of the 3-tone pattern (i.e., 18 tones) were required between trials. (There was no maximum.)

Prior to the test phase, there was a training phase designed to familiarize the infants with the contingency between head turning and

animated toy reinforcement, using an easier task (200-msec gaps). There were no control trials during training, but otherwise the procedure was identical to that used during the test phase. The infants were required to reach a training criterion of four consecutive correct responses within 20 trials in order to proceed to the test phase.

The adults were tested in essentially the same procedure with the same training criterion but raised their hand rather than turned their head when they detected a gap increase. It is likely that the toys were reinforcing for adults as well; at a minimum, they provided feedback. The adults completed a questionnaire outlining their musical experience.

Results and Discussion

Since the infants and the adults were tested with different silent gap durations, absolute performance could not be compared. Separate analyses of variance (ANOVAs) on the percentage of correct responses (the proportion out of 10 trials × 100), with gap location (SgSL, SSgL, SSLg, control) as a within-subjects factor and starting tone (SSL, SLS, LSS) as a between-subjects factor, were conducted for the infants and the adults. For both groups, only gap location was significant [$F(3,81) = 24.37$, $p < .0001$, for infants; $F(3,81) = 146.19$, $p < .0001$, for adults; see Figure 2].

Since it was hypothesized that the participants would group as [SSL]—that is, hear a segment boundary after the long complex tones—better performance was expected on trials SgSL and SSgL (within-group gaps) than on SSLg (between-group gap). This is exactly what was found. Tukey HSD tests were used to compare performance across the four levels of gap location. With four means to compare and the MS_e (infants: 2.20; adults: 3.41) and $df(81)$ from the ANOVAs above, critical difference values ($q\sqrt{MS_e/n}$) for $\alpha = .05$ were 1.25 for adults and 1.01 for infants. Consistent with expectations, for both the infants and the adults, performance on SgSL (infants: mean = 54.0%, $SE = 2.7$; adults: mean = 90.9%, $SE = 3.2$) and SSgL (infants: mean = 46.6%, $SE = 3.9$; adults: mean = 77.7%, $SE = 5.0$) did not differ significantly (although this effect approached conventional levels of significance for the adults, $p < .06$). On the other hand, performance on SgSL was significantly better than performance on SSLg (infants: mean = 35.7%, $SE = 2.9$; adults: mean = 35.0%, $SE = 5.5$; $p < .0002$ for infants, $p < .0002$ for adults), and performance on SSgL was significantly better than performance on SSLg ($p < .03$ for infants; $p < .0002$ for adults). The percentage of correct responses on each of the three gap location trials was significantly greater than the percentage of correct responses on control trials (infants: mean = 23.3%, $SE = 2.3$; adults: mean = 1.0%, $SE = 0.7$; all $ps < .01$) for both the infants and the adults, simply indicating that the infants and the adults could hear all gap increases at above-chance levels. The nonsignificant trend for both the infants and the adults to perform better on SgSL than on SSgL (see Figure 2) might reflect the fact that it is easier to temporally encode two similar sounds than two different sounds (Thorpe & Trehub, 1992).

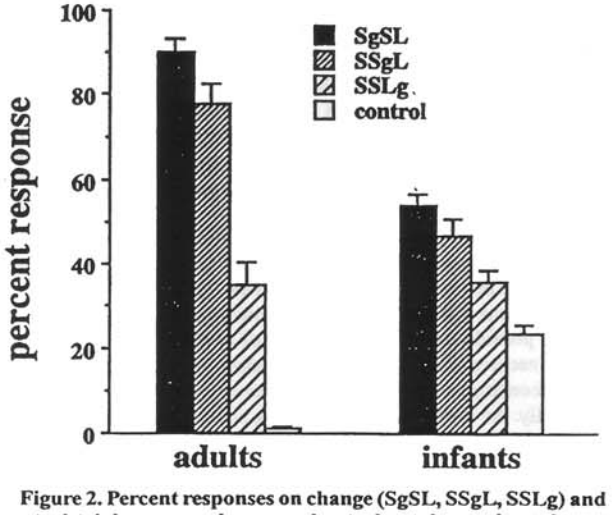


Figure 2. Percent responses on change (SgSL, SSgL, SSLg) and control trials averaged across short–short–long, short–long–short, and long–short–short conditions for infants and adults (Experiment 1; g stands for gap). Error bars represent the standard error of the mean.

The number of years of musical training was not significantly correlated with overall performance ($p > .6$). Although there was not a very large range for the musical training present, there is no indication that amount of musical training affects performance on this task.

Lower response rates to gaps in location SSLg than in SgSL and SSgL suggest that an increase in the duration of a component of a repeating complex may have marked the ends of groups for both the infants and the adults. However, this conclusion depends on the assumption that sensitivity to increases in the 200-msec gap between tones was only a function of perceptual grouping, and not of the duration of the markers (i.e., the tones defining the beginning and the end of the silence) in the absence of a repeating pattern. Abel (1972) measured adults' 75% thresholds for detecting changes in standard silent gaps bounded by Gaussian noise bursts of 10 or 300 msec. The standard gaps ranged from less than 1 msec to over 600 msec. Gaps of 200 msec were not explicitly tested. However, the average Weber fraction for a marker duration of 300 msec with standard gaps of 160 or 320 msec was .26. This would predict that, for 200-msec gaps, the 75% correct threshold for detecting a change in gap duration would be 52 msec. For 10-msec marker durations matched to the 300-msec markers in amplitude, the Weber fraction was very similar (.28), which leads to a prediction of 75% correct thresholds of 56 msec for 200-msec gaps. However, for a 10-msec marker matched in overall energy (intensity × time), the Weber fraction was .21, which leads to a 75% correct prediction of 42 msec for 200-msec gaps. These data suggest that changing marker duration thirtyfold (from 10 to 300 msec) increases 75% thresholds with 200-msec markers by, at most, 10 msec. The markers in Experiment 1 of the present study changed only threefold, from 200 to 600 msec. Thus, it seems unlikely that effects of the marker durations in isolation could be

responsible for the results. It is also worth pointing out that the adults in the present experiment appear to be performing considerably better at detecting gaps in the context of the rhythmic pattern (they are over 90% correct at detecting gap SgSL when g is 30 msec on a base of 200 msec), as compared with the subjects in Abel's study, whose 75% correct thresholds for 200-msec gaps in isolation would be somewhere between 42 and 56 msec.

Although the context of a rhythmic pattern may improve performance to the extent that effects of markers in isolation are not contributing factors, and Abel's (1972) data suggest that it is likely that the differences in gap detection between 200- and 600-msec markers are very small, one must be careful in extrapolating to untested marker durations. Experiment 1A was conducted in order to directly test the effects of the 200- and 600-msec markers on gap detection in isolation in adults.

EXPERIMENT 1A

Method

Participants. Ten adults (mean age = 19 years; 2 male and 8 female) with no known hearing loss participated. None was a professional musician (5 had between 0 and 5 years of music lessons; 5 had over 5 years). Number of years of music lessons was not correlated with performance on any of the dependent measures.

Stimuli. The 200-msec (S) and 600-msec (L) complex tones from Experiment 1 were used to mark the beginnings and the ends of gaps. On each trial, the adults were to judge whether two silent gaps, each surrounded by tones, were the same or different. There were four kinds of trials, each with different markers (i.e., tone durations) at the beginning and the end of the gaps: two short tones (SS), a short and a long tone (SL), a long and a short tone (LS), and two long tones (LL). The two gaps to be compared within a trial always had the same markers. The two gap-with-markers stimuli within each trial were separated by 1,600 msec. In order to match the gap detection task of Experiment 1, the initial gap of every trial was 200 msec. On *same* trials, the second gap was also 200 msec (e.g., for SS same trials, participants heard SS–SS), but on *different* trials it was 200 + g msec, where g was 100 msec during training and 30 msec during testing, as in Experiment 1 (e.g., for SS different trials, participants heard SS–SgS). Thus, the four types of same trials were SS–SS, SL–SL, LS–LS, and LL–LL. The four types of different trials were SS–SgS, SL–SgL, LS–LgS, and LL–LgL.

Apparatus and Procedure. The apparatus was identical to that in Experiment 1. The adults were tested in the sound-attenuating chamber. They were initially given 20 training trials. Each adult then completed 80 test trials, 20 each with SS, SL, LS, and LL markers. Half of the trials were *same* trials, and half were *different*. The subjects initiated each trial by pressing a button on the button box. They entered their response by pressing either the "same" or the "different" button on the box.

Results and Discussion

The percent correct was calculated for each of the SS, SL, LS, and LL conditions, in order to facilitate comparisons with Experiment 1. However, d' analyses were also conducted, and the conclusions are exactly the same whether percent correct or d' is used. A repeated measures ANOVA did not reveal a significant difference across the SS, SL, LS, and LL conditions ($p > .2$). In fact, this was not surprising, in light of the fact that overall performance was only marginally different from chance levels [$t(9) = 1.76$, $p = .06$; $M = 53.5\%$ correct, $SD = 6.3$].

The first conclusion is that gap detection performance is superior in a rhythmic pattern context than in isolation. Presumably, the pattern allows more precise prediction as to when the next complex tone should occur. This is an interesting phenomenon, which deserves more investigation, although it is beyond the scope of the present paper.

The second conclusion, and the one most directly relevant to this paper, is that the differences across gap location found in Experiment 1 were not due to differential effects of the markers independent from the repeating pattern context, because a 30-msec increment in gap size between two markers presented in isolation cannot be detected. We are confident in concluding that longer duration in the context of a repeating tone pattern is used as a cue for perceiving the end of a group.

EXPERIMENT 2

The results of Experiment 1 were consistent with the hypothesis that both infants and adults use longer duration elements to mark the ends of groups. As outlined in the introduction, intensity is another component of prosody and rhythm that could be used as an acoustic cue to perceptual grouping. Experiment 2 addresses the question of whether an increase in intensity marks the beginnings of groups.

Method

Participants. Thirty 8-month-old full-term, healthy infants (mean age = 8 months 13 days; 14 male, 16 female) were tested (10 per condition). One further infant was eliminated for failing to meet the training criterion. Thirty adults (mean age = 22 years; 9 male, 21 female) with no known hearing loss were also tested. Again, none was a professional or a serious amateur musician (19 had between 0 and 5 years of music lessons; 11 had over 5 years). None of the participants from Experiments 1 or 1A participated in Experiment 2.

Stimuli. The stimuli were identical to those of Experiment 1, with the following exceptions. All the complex tones were 200 msec, including 10-msec rise and decay times, and were separated by 200-msec silent intervals. Every third tone was 10 dB more intense than the other two (65 vs. 55 dB[A]). Six-month-old infants' thresholds for intensity discrimination are about 6 dB for 1000-Hz pure tones (Sinnott & Aslin, 1985), but presumably lower for complex tones. Bull et al. (1984) found that infants of between 5 and 11 months could perceive intensity differences as small as 2 dB in vowels. Thus the 10-dB difference in the present context should be easily discriminable to both infants and adults.

There were three conditions, each starting on a different tone of the pattern, named LQQ (loud–quiet–quiet–loud–quiet–quiet ...), QLQ (quiet–loud–quiet–quiet–loud–quiet ...), and QQL (quiet–quiet–loud–quiet–quiet–loud ...).

Again, there were four types of trials—a control trial involving no gap increase and three types of change trials: LgQQ (gap inserted between loud and quiet); LQgQ (gap inserted between quiet and quiet); and LQQg (gap inserted between quiet and loud; see Figure 1b). It was expected that the subjects would group the sequences as [Loud–Quiet–Quiet] and that performance would be worse on LQQg than on either LgQQ or LQgQ.

Apparatus and Procedure. The apparatus and procedure were identical to those of Experiment 1, except that the infants and the

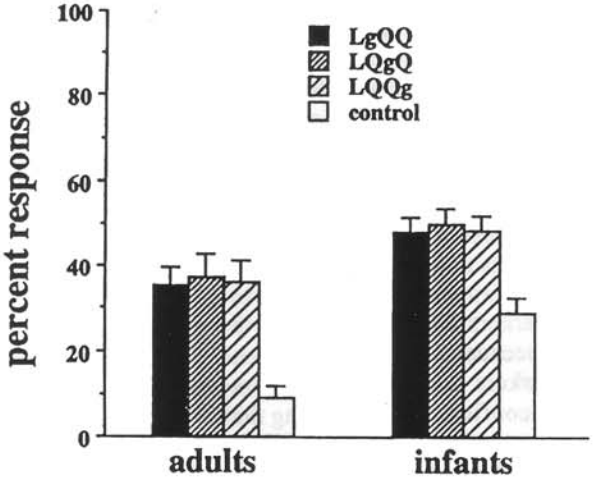


Figure 3. Percent responses on change (LgQQ, LQgQ, LQQg) and control trials averaged across loud–quiet–quiet, quiet–loud–quiet, and quiet–quiet–loud conditions for infants and adults (Experiment 2; g stands for gap). Error bars represent the standard error of the mean.

adults were tested in the LQQ, QLQ, and QQL conditions. Again, there were 40 trials in each condition, 10 each of LgQQ, LQgQ, and LQQg and 10 control trials.

Results and Discussion

The same analyses were performed as those in Experiment 1 (see Figure 3). ANOVAs, with gap location (LgQQ, LQgQ, LQQg, control) as a within-subjects factor and starting note (LQQ, QLQ, QQL) as a between-subjects factor, revealed an effect of gap location for infants [$F(3,81) = 10.94$, $p < .0001$] and for adults [$F(3,81) = 17.17$, $p < .0001$]. Tukey HSD tests were used to compare performance across the four levels of gap location. With four means to compare and the MS_e (infants: 2.70; adults: 3.20) and $df(81)$ from the ANOVAs above, critical difference values ($q\sqrt{MS_e/n}$) for $\alpha = .05$ were 1.11 for infants and 1.21 for adults. For both the infants and the adults, the effect was due entirely to fewer responses (i.e., false alarms) occurring on control trials (infants: mean = 29.0%, $SE = 3.7$; adults: mean = 9.3%, $SE = 2.5$) than on any of the three types of gap trials (all $ps < .0003$). There were no significant differences between any of the three types of gap trials—LgQQ (infants: mean = 48.0%, $SE = 3.5$; adults: mean = 35.3%, $SE = 4.5$), LQgQ (infants: mean = 50.0%, $SE = 3.4$%; adults: mean = 37.3%, $SE = 5.5$), and LQQg (infants: mean = 48.3%, $SE = 3.8$; adults: mean = 36.3%, $SE = 4.9$; all $ps > .9$). Thus, there was no evidence that increasing the intensity of every third complex tone leads to systematic grouping of any kind in infants.

For adults, there was also an interaction between gap and starting note [$F(6,81) = 4.81$, $p < .0004$]. To examine this effect, separate Tukey tests were conducted for each of the three starting note conditions. Recall that, once the experiment began, the three conditions were identical, and the pattern repeated continuously throughout the test session. The conditions differed only in where the pattern began at the very beginning of the test session. Although the differences between means were not always significant, the trends were consistent with the adults' tending to group according to the first three complex tones they heard in the experiment. As can be seen in Figure 4, when the experiment began with LQQ . . . , performance was lowest on LQQg, suggesting that the adults were grouping as [LQQ]. On the other hand, when the experiment began with QLQ . . . , performance was lowest on LQgQ, suggesting that the adults were grouping as [QLQ], and when the experiment began with QQL . . . , performance was lowest on LgQQ, suggesting that the adults were grouping as [QQL]. Thus, there was no evidence that adults group in such a way that louder tones mark the beginnings of groups; rather, there were trends suggesting that the very first pattern heard biases the perceived grouping.

GENERAL DISCUSSION

The results are consistent with the hypothesis that both 8-month-olds and adults use an increase in duration to mark the ends of groups and that this effect is independent of the effects of tone duration in the absence of a repeating pattern. However, intensity did not mark the beginnings of groups at either age; in fact, no systematic effect of intensity on segmentation was found. This is particularly interesting in light of the fact that longer elements are most often described as sounding stressed and sounding louder, even when there is no difference in intensity (Fry, 1955; Woodrow, 1909). It is possible that intensity may play a moderating role, interacting with the effects of other factors, such as duration and pitch.

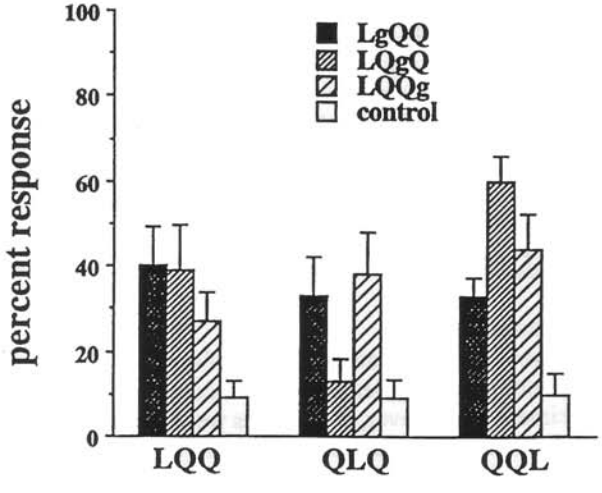


Figure 4. Percent responses for adults in Experiment 2 on change and control trials for each starting-note condition—loud–quiet–quiet (LQQ), quiet–loud–quiet (QLQ), and quiet–quiet–loud (QQL)—separately.

Stressed syllables in speech and stressed tones in music tend to be both longer and more intense. Interestingly, however, a number of studies have shown that intensity has a much smaller effect on perceived stress (Fry, 1955) and plays a much smaller role in determining word (Nakatani & Schaffer, 1978) and phrase boundaries (Streeter, 1978) than does duration. Streeter found evidence that intensity primarily moderated the effect of duration but had little independent effect. In music performance, there is evidence that duration variation plays a more important role than intensity variation (Drake, 1993; Drake, Dowling, & Palmer, 1991), and 5- to 8-year-old children easily reproduce the duration pattern of a musical rhythm, but not the intensity pattern (Gérard & Drake, 1990). Thus, the small number of relevant studies are consistent with the interpretation that duration plays a more important role than intensity in the segmentation of both syllable streams and tone streams.

Phrase-final lengthening is found across most languages and also appears to be a characteristic of musical compositions. The results of the present study indicate that both infants and adults use lengthened elements to signal the ends of groups in auditory complex tone patterns that are neither speech nor music. This leads to a number of questions concerning the development of sensitivity to segmentation and the developmental relation between linguistic and musical segmentation. By 8 months, infants have had considerable exposure to language and music, so it remains unclear as to when infants learn to use phrase-final lengthening for segmentation.

It is interesting that phrase-final lengthening appears to be a cue to segmentation across various types of auditory sequences, raising the possibility that there is a general segmentation mechanism that may be employed in both linguistic and musical analysis. One possibility is that, in development, there is initially a single segmentation mechanism but that it becomes differentiated as infants acquire language-specific and musical-system-specific knowledge. There are few studies of the development of sensitivity to phrase structure. Using natural stimuli, Jusczyk et al. (1992) found that 9-month-old, but not 6-month-old, infants were sensitive to linguistic phrasal units, whereas Hirsh-Pasek et al. (1987) found that infants as young as 7 months were sensitive to clausal units. Using a simple Mozart minuet, Jusczyk and Krumhansl (1993) found that 4½-month-olds were sensitive to musical phrase structure. Although this seems to suggest different time courses of development for linguistic and musical stress, it is difficult to compare across these studies, because of the presence of multiple cues to segmentation. The problem is compounded by variation in the extent of pitch, intensity, and duration cues and by the presence of different cues, such as phonotactic cues in the linguistic case and octaves at boundaries in the musical case. Certainly, the specific role of preboundary lengthening in infants' perceptions of phrase endings in these studies is not known. Thus, in order to examine the relation between musical and linguistic segmentation, it will be necessary to conduct a series of studies in which the various cues to segmentation are tightly controlled. By testing whether infants of various ages show different patterns of segmentation across linguistic, musical, and tone sequences, these questions can be addressed.

REFERENCES

Abel, S. M. (1972). Discrimination of temporal gaps. *Journal of the Acoustical Society of America*, **52**, 519-524.

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

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, **6**, 145-238.

Boltz, M. G. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception & Psychophysics*, **53**, 585-600.

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.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic Press.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant directed speech in the first month after birth. *Child Development*, **61**, 1584-1595.

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

Cooper, W. E., & Sorensen, J. M. (1977). Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America*, **62**, 683-692.

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, **266**, 718-719.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, **28**, 381-389.

Drake, C. (1993). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society*, **31**, 107-110.

Drake, C., Dowling, W. J., & Palmer, C. (1991). Accent structures in the reproduction of simple tunes by children and adult pianists. *Music Perception*, **8**, 315-334.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception*, **10**, 343-378.

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

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.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preferences for motherese speech. *Infant Behavior & Development*, **10**, 279-293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, **27**, 209-221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, **20**, 104-113.

Fowler, C. A., Smith, M. R., & Tassinary, L. G. (1985). *Journal of the Acoustical Society of America*, **79**, 814-825.

Fraisse, P. (1956). *Les structures rhythmiques* [Rhythmic structures]. Louvain: Publications Universitaires de Louvain.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149-180). New York: Academic Press.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, **27**, 765-768.

Gérard, C., & Drake, C. (1990). The inability of young children to reproduce intensity differences in musical rhythms. *Perception & Psychophysics*, **48**, 91-101.

Gerken, L. A., Jusczyk, P. W., & Mandel, D. R. (1994). When

prosody fails to cue syntactic structure: Nine-month-olds' sensitivity to phonological vs. syntactic phrases. *Cognition*, **51**, 237-265.

GLEITMAN, L., GLEITMAN, H., LANDAU, B., & WANNER, E. (1988). Where the learning begins: Initial representations for language learning. In R. Newmeyer (Ed.), *The Cambridge linguistic survey* (pp. 150-193). Cambridge, MA: Harvard University Press.

HANDEL, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

HIRSH-PASEK, K., KEMLER NELSON, D. G., JUSCZYK, P. W., WRIGHT-CASSIDY, K., DRUSS, B., & KENNEDY, L. (1987). Clauses are perceptual units for young infants. *Cognition*, **26**, 269-286.

JUSCZYK, P. W., CUTLER, A., & REDANZ, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, **64**, 675-687.

JUSCZYK, P. W., HIRSH-PASEK, K., KEMLER NELSON, D. G., KENNEDY, L., WOODWARD, A., & PIWOZ, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, **24**, 252-293.

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

JUSCZYK, P. W., & THOMPSON, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics*, **23**, 105-109.

KIDD, G., BOLTZ, M., & JONES, M. R. (1984). Some effects of rhythmic context on melody recognition. *American Journal of Psychology*, **97**, 153-173.

KLATT, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208-1221.

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

MARTIN, J. G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning & Verbal Behavior*, **9**, 75-78.

MEHLER, J., JUSCZYK, P. W., LAMBERTZ, G., HALSTED, N., BERTONCINI, J., & AMIEL-TISON, C. (1988). A precursor of language acquisition in young infants. *Cognition*, **29**, 143-178.

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

MORGAN, J. L. (1994). Converging measure of speech segmentation in preverbal infants. *Infant Behavior & Development*, **17**, 389-403.

MORGAN, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory & Language*, **35**, 666-688.

MORGAN, J. L., & DEMUTH, K. (1996). *Signal to syntax*. Hillsdale, NJ: Erlbaum.

MORRONGIELLO, B. A., KULIG, J. W., & CLIFTON, R. K. (1984). Developmental changes in auditory temporal perception. *Child Development*, **55**, 461-471.

MORRONGIELLO, B. A., & TREHUB, S. E. (1987). Age-related changes in auditory temporal perception. *Journal of Experimental Child Psychology*, **44**, 413-426.

NAKATANI, L. H., & SCHAFFER, J. A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, **63**, 234-245.

POVEL, D.-J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 3-18.

POVEL, D.-J., & ESSENS, P. (1985). Perception of temporal patterns. *Music Perception*, **2**, 411-440.

POVEL, D.-J., & OKKERMAN, H. (1981). Accents in equitone sequences. *Perception & Psychophysics*, **30**, 565-572.

PREUSSER, D., GARNER, W. R., & GOTTWARD, R. L. (1970). Perceptual organization of two-element temporal patterns as a function of their component one-element patterns. *American Journal of Psychology*, **83**, 151-170.

ROYER, F. L., & GARNER, W. R. (1966). Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception & Psychophysics*, **1**, 41-47.

SAFFRAN, J. R., ASLIN, R. H., & NEWPORT, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926-1928.

SCOTT, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, **71**, 996-1007.

SINNOTT, J. M., & ASLIN, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *JASA*, **78**, 1986-1992.

SPRING, D. R., & DALE, P. S. (1977). Discrimination of linguistic stress in early infancy. *Journal of Speech & Hearing Research*, **20**, 224-231.

STREETER, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, **64**, 1582-1592.

THORPE, L. A., & TREHUB, S. E. (1992). Duration illusion and auditory grouping in infancy. *Developmental Psychology*, **25**, 112-127.

TODD, N. (1985). A model of expressive timing in tonal music. *Music Perception*, **3**, 33-58.

TRAINOR, L. J. (1996). Infant preferences for infant-directed versus non-infant-directed playsongs and lullabies. *Infant Behavior & Development*, **19**, 83-92.

TRAINOR, L. J., CLARK, E. D., HUNTLEY, A., & ADAMS, B. (1997). Comparisons of infant-directed and non-infant directed singing: A search for the acoustic basis of infant preferences. *Infant Behavior & Development*, **20**, 383-396.

TRAINOR, L. J., & TREHUB, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 394-402.

TREHUB, S. E., SCHNEIDER, B. A., & HENDERSON, J. L. (1995). Gap detection in infants, children, and adults. *Journal of the Acoustical Society of America*, **98**, 2532-2541.

TREHUB, S. E., & THORPE, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology*, **43**, 217-229.

TREHUB, S. E., & TRAINOR, L. J. (1993). Listening strategies in infancy: The roots of music and language development. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: Cognitive perspectives on human audition* (pp. 278-327). Amsterdam: Elsevier.

TREHUB, S. E., TRAINOR, L. J., & UNYK, A. M. (1993). Music and speech processing in the first year of life. *Advances in Child Development & Behavior*, **24**, 1-35.

WERNER, L. A., CAMERON, M., HALPIN, C. F., BENSON SPETNER, N., & GILLENWATER, J. M. (1992). Infant auditory temporal acuity: Gap detection. *Child Development*, **63**, 260-272.

WOODROW, H. (1909). A quantitative study of rhythm. *Archives of Psychology*, **14**, 1-66.

WOODROW, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224-1236). New York: Wiley.

NOTE

1. Temporal processing in infants is relatively good: By 6 months of age, duration discrimination thresholds (Morrongiello & Trehub, 1987) and the asynchrony needed for the precedence effect (Morrongiello, Kulig, & Clifton, 1984) appear to be about double those of adults. Infant gap detection thresholds appear to vary as a function of the method used. With short Gaussian-modulated tone pips, 6-month-old infants' thresholds are around 12 msec (Trehub, Schneider, & Henderson, 1995); with low-pass continuous broadband noise, however, they are around 40–60 msec, depending on the frequency cutoff (Werner, Cameron, Halpin, Benson Spetner, & Gillenwater, 1992).

(Manuscript received May 27, 1997;
revision accepted for publication November 11, 1998.)