

Organization and discrimination of repeating sound sequences by newborn infants^{a)}

Stephen McAdams^{b)}

Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes, EPHE, 28 rue Serpente, F-75006 Paris, France and Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), 1 place Igor-Stravinsky, F-75004 Paris, France

Josiane Bertoncini

Laboratoire de Sciences Cognitives et Psycholinguistique (CNRS), Ecole des Hautes Etudes en Sciences Sociales, 54 bd Raspail, F-76270 Paris 06, France

(Received 18 September 1996; accepted for publication 28 July 1997)

A study was conducted to determine whether newborn infants organize auditory streams in a manner similar to that of adults. A series of three experiments investigated the ability of 3- to 4-day-old infants to discriminate repeated rising and falling four-tone sequences in two configurations of source timbre and spatial position. It was hypothesized that if the sequences were organized into two auditory streams on the basis of timbre and spatial position, one of the configurations should be discriminable from its reversal while the other should not. The sequences were tested with different pitch and temporal intervals separating the tones. Sequences were discriminated for the first configuration by adults at both fast tempo/small interval and slow tempo/large interval combinations, while only the latter was discriminated by newborns as measured with a non-nutritive high-amplitude sucking paradigm. Neither adults nor infants could discriminate the sequence reversals for the second configuration. The results suggest that newborn infants organize auditory streams on the basis of source timbre and/or spatial position. They also suggest that newborns have limits in temporal and/or pitch resolution when discriminating tone sequences. © 1997 Acoustical Society of America. [S0001-4966(97)03411-5]

PACS numbers: 43.66.Mk, 43.66.Jh, 43.66.Qp [WJ]

INTRODUCTION

The acoustic environment is composed of numerous sources of sound, and one of the important tasks for any animal is to be able to perceive them separately and to perform actions with respect to them. This ability may be considered to involve the building up of a veridical mental representation of the sources of sound present in the environment, a representation that is then used to plan appropriate action. One class of processes that seems to be involved in this kind of perceptual organization is the connecting together through time of individual sound events that are emitted by the same source, a process called sequential auditory organization or auditory streaming (Bregman and Campbell, 1971; McAdams and Bregman, 1979; Bregman, 1990). Several acoustic factors have been shown to play a role in streaming, such as spectral discontinuity (Bregman *et al.*, 1990), intensity discontinuity (van Noorden, 1977), and spatial discontinuity (Hartmann and Johnson, 1991). That is, a sequence composed of sounds that are more or less similar with respect to these three dimensions tends to be heard as a single sound stream, whereas a sequence that alternates between two regions of relatively distant values along one or more of these dimensions tends to be heard as two streams.

For most sound sources in the everyday world, the various sound dimensions give rise to unambiguous organizations into sound objects and allow a clear understanding of their behavior through time. In addition, this organization appears to take place automatically without much conscious or deliberate intervention on the part of the organism. An important question that arises in the face of this evidence is whether the processes of sequential auditory organization are part of our innate perceptual equipment (present at birth or maturing within the first few weeks of life), or whether they are acquired through experience in the acoustic world. From birth, infants have to deal with an acoustic environment in which many simultaneous sources of sound can compete. Moreover, it is from such noisy situations that infants must correctly extract highly relevant acoustic information such as their mother's voice among others speaking her native language. Surprisingly, there is almost no research on infants' capacity to organize their acoustic world in terms of separate sources, or to perceive auditory streams. Auditory development has received much interest (e.g., Trehub and Trainor, 1993; Werner and Rubel, 1992), particularly in terms of the development of speech perception (e.g., Jusczyk, 1997), but very few studies have addressed the problem of how infants perceive complex sounds as coherent units, similarly to what is called object perception in visual cognition (cf. Spelke, 1990; Spelke *et al.*, 1992). Although scarce, the data available seem nonetheless to favor the existence of unlearned, basic mechanisms engaged in streaming very early in life.

^{a)}Portions of these results were presented at the 120th Meeting of the Acoustical Society of America, San Diego [McAdams *et al.*, *J. Acoust. Soc. Am. Suppl.* 1 **88**, S91 (1990)].

^{b)}Electronic mail: smc@ircam.fr

Some arguments in this direction have been proposed by Bregman (1990).

The results of two previous studies (Demany, 1982; Fassbender, 1993) are consistent with such an assumption. They have addressed the question directly by testing young infants in auditory conditions that lead adults to perceive sound sequences as organized into two different streams. Demany (1982) demonstrated with a visual fixation procedure that 1.5- to 3-month-old infants organize sound sequences according to spectral proximity (frequency region of pure tones). Fassbender (1993) further demonstrated with a non-nutritive sucking paradigm that 2- to 5.5-month-old infants organize sound sequences on the basis of frequency proximity, intensity similarity, and spectral similarity. From these results it appears that stream segregation processes are operative in the very first weeks of life, at least when streams of fairly simple sounds are separated according to a number of acoustic dimensions known to be effective in promoting segregation in adults. Other properties of sound sources that are powerful cues for detecting and recognizing auditory events in the natural environment, such as correlated multidimensional variation of source timbre and spatial location, have not yet been tested with infants. Thus the present study aimed to extend previous research in two ways. First, we tested *newborns* to verify whether stream segregation is part of the perceptual apparatus that first encounters the acoustic world. And second, but more importantly, we tested their capacity to use the *timbre and spatial position* of naturalistic sounds to perceptually organize sound sequences in terms of two distinct auditory entities. The timbres of the sound sources (vibraphone and trumpet sounds) as well as their spatial locations were chosen to be sure that the differences were easily discriminable by young infants (see Clifton, 1992, for a review on sound localization in infants, and Trehub *et al.*, 1990, concerning processing of timbre). In this study, we do not question infants' ability to distinguish vibraphone from trumpet or sounds coming from the right or left. We assume that this ability is part of the newborn's perceptual repertoire. Our question is whether such differences in timbre and spatial position are automatically used by newborns' to perceive sound sequences as originating from two individual sources.

A melody discrimination paradigm similar to the one used by Demany (1982) was employed to probe stream formation in both newborn infants and adults. A test of infants' ability to discriminate a melody from its retrograde (or reverse order of the tones) was first performed in order to find stimulus conditions under which they could perform the discrimination using a non-nutritive sucking paradigm (experiments 1 and 2). Similar conditions were presented to adults for comparison (experiment 4). Then the discriminable melody patterns were presented under conditions that adults perceive as two separate sound streams organized on the basis of timbre and spatial position (experiment 3 for newborns, experiment 4 for adults). If the latter sequence is organized by newborns on the basis of timbral and spatial similarity, then the sequence discrimination demonstrated in experiments 1 and 2 should fail in experiment 3.

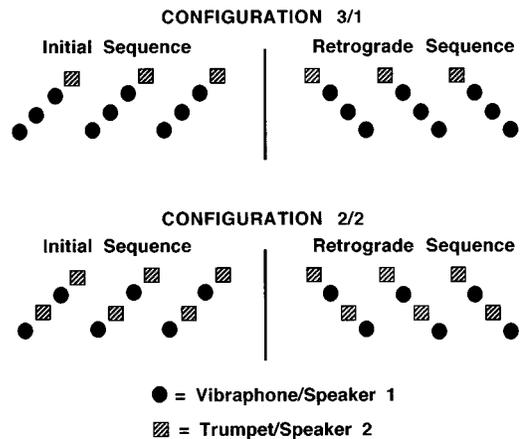


FIG. 1. Three cycles of each of the four-tone repeating stimulus patterns used in the melodic discrimination paradigm are shown. Pitch corresponds to the vertical dimension and time to the horizontal dimension. Timbre and spatial location of the sound sources are shown by the form and shading of the events. Speakers were placed to the right and left of the infant's head.

I. GENERAL METHOD

A. Stimuli

In the interest of giving the infants as many cues as possible for organizing the streams in terms of complex sound sources, it was decided to configure sequences on the basis of two cues, operating in perfect conjunction. Thus two configurations of timbre and spatial configuration were applied to repeating melodic patterns. In these configurations, one speaker was present on each side of the baby's head. Two synthetic timbres created on a Yamaha digital synthesizer were used that simulated a slightly inharmonic, metallic percussion instrument (vibraphone) and a brass wind instrument (trumpet). The instruments that were simulated by the synthesized sound differ in both the resonator (bar versus air column) and the way the resonator is excited (impulsively versus continuously). These specific timbres were previously found by McAdams *et al.* (1995) to be perceived by adult subjects as very dissimilar along several perceptual dimensions (attack time, spectral distribution, and degree of spectral evolution). The timbres were equalized for loudness across the pitch range used in this experiment by author SM. For a given subject, each instrument always appeared in the same speaker and always played the same pitches.

Repeating melodic patterns of four ascending or descending pitches were used (Fig. 1). The experiment was based on the ability to discriminate an initial melodic contour¹ (rising or falling) from its retrograde (falling or rising, respectively). Trehub *et al.* (1987) and Ferland and Mendelson (1989) have demonstrated that 9- to 11-month-old infants can discriminate and categorize these kinds of contours, although, to our knowledge, no data are available on this capacity in newborns.

The first such configuration is labeled 3/1 due to the fact that three tones are played by the vibraphone in one speaker and one tone is played by the trumpet in the other speaker. Under the hypothesis that the sequences would be organized into perceptual streams on the basis of timbre and/or spatial position of the tone events, it was expected that two streams

would arise from these sequences: one with three pitches in a rising or falling pattern, and another with a single repeating pitch. Our experimental hypothesis was that the initial and retrograde patterns would be discriminable due to the difference of the melodic contours in the vibraphone stream. In fact, for this configuration, even if the subject paid no attention to timbre and spatial position, the sequences should be discriminable if the melodic contour has been stored and can be compared across sequences. The reason that two timbres and spatial positions are included in this sequence is to avoid confounding configurational differences between this sequence and the second one to be described below with differences in the number of timbres and spatial positions present in each sequence.

The second configuration is labeled 2/2 since the four pitches are split into two interleaved groups, with pitches 1 and 3 being assigned to the vibraphone and pitches 2 and 4 to the trumpet. Under the hypothesis that the sequences are organized on the basis of these parameters, two perceptual streams should result, each one consisting of a pattern alternating between two pitches. Our experimental hypothesis was that for this configuration, the initial and retrograde patterns would not be discriminable since each stream is perfectly symmetric in terms of its melodic pattern. Therefore, if the newborns can discriminate the 3/1 configuration and cannot discriminate the 2/2 configuration, this result would allow us to argue that they organize sound sequences on the basis of timbre and/or spatial position of the tone events.

Stimuli were synthesized on a Yamaha TX802 FM Tone Generator controlled by a Macintosh computer. The musical instrument simulations were developed by Wessel *et al.* (1987). The stimuli were recorded on a stereo cassette tape.

B. Procedure

An habituation/dishabituation paradigm was used in which sound presentation was controlled by the infant's sucking on a non-nutritive pacifier (Jusczyk, 1985; Floccia *et al.*, 1997). Each time the amplitude of the infant's suck exceeded a fixed threshold, several cycles of a repeating four-tone sequence were played. It was decided to present a minimum of three to four cycles in order to ensure that stream segregation occurred, since Bregman (1978) has shown that this process takes time. A computer program recorded the sucking rate in 1-min periods. The habituation point was considered to have occurred when the sucking rate fell by at least one-third for 2 consecutive min compared to the 1-min period immediately preceding these 2 min. At this point, the stimulus sequence was changed for the experimental group and remained identical for the control group. The infants were considered to have discriminated the sequence change if the difference in mean sucking rate between the 1-min periods before and after habituation was significantly greater for the experimental group than for the control group.

The tapes were played over loudspeakers placed on either side of the infant's head. They were elevated by about 20 cm and formed an azimuthal angle of approximately 120° that was bisected by the orientation of the infant's head. The stimuli were presented at an A-weighted level of about 70

dB in a single-walled sound isolation chamber. The frequency components of the complex sounds used would generally be well above pure-tone thresholds in newborns when presented at this level (cf. Werner, 1992).

Babies were rejected during experimentation if they refused to suck on the pacifier, failed to reach the habituation criterion within 15 min, lost the pacifier during the 3 min prior to or following habituation, fell asleep, cried, or became agitated.

C. Subjects

Subjects in experiments 1–3 were newborn infants of 3 to 4 days of age who were selected to participate in the experiment on the basis of their health record during pregnancy, delivery, and the 3 or 4 days after birth. The selection criteria included the following: (1) their weight at birth had to exceed 2.8 kg; (2) the gestation period had to be at least 38 weeks; (3) their APGAR score (a general measure of health and responsiveness) had to attain the value of 10 at least by the fifth minute following birth; and (4) they had to be in good health at the time of testing. After having obtained the permission of the parents, infants were brought to the experimental situation about 2 1/2 h after their last feeding and 30 min to 1 h before their next one. The newborn infants were tested at the Baudelocque Maternity Hospital in Paris.

II. EXPERIMENT 1

Experiment 1 tested the hypothesis that infants should be able to discriminate the original 3/1 configuration from its retrograde version.

A. Method

Stimulus sequences consisted of a repeating four-note pitch pattern (C4–D4–E4–F#4, or its retrograde, F#4–E4–D4–C4) presented at a rate of 10 tones/s (an inter-tone onset interval of 100 ms). Each tone had a duration of 85 ms. Pitches were assigned to either the left or right channel in order to effect spatial separation and to one of two instruments in order to effect a separation based on source timbre (Fig. 1). In these sequences, the three lowest pitches (C, D, E) were assigned to one channel and presented with the vibraphone timbre. The highest pitch (F#) was presented with the trumpet timbre in the other channel. The side of presentation of the timbres was counterbalanced across subjects. The initial pattern was either the ascending or descending contour, each being presented to an equal number of subjects.

The subjects were randomly assigned to one of two independent groups: Experimental, with contour change at habituation, and Control, with no contour change. Twenty subjects completed the experiment in each group for a total of 40 subjects. Data for 32 additional subjects were rejected: 17 did not habituate within 15 min, 4 produced insufficient or irregular sucking or spat out the pacifier, 3 fell asleep, 7 cried or became agitated, and 1 was removed due to experimenter error.

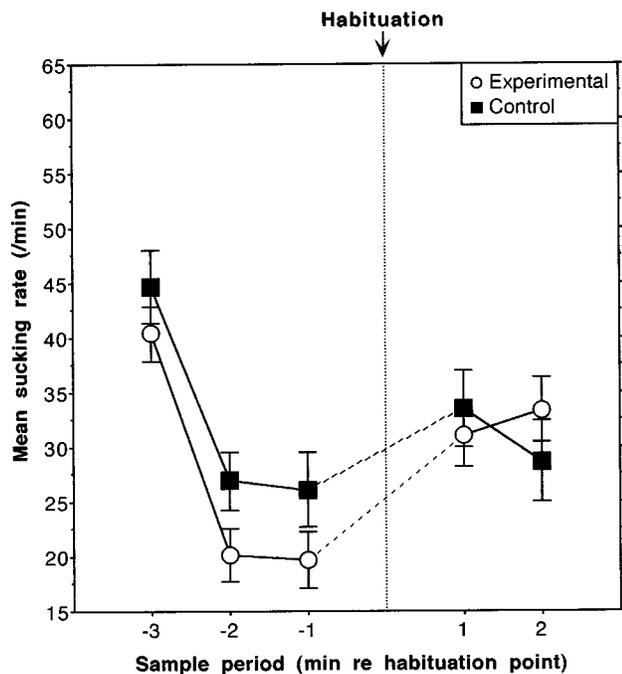


FIG. 2. Results for experiment 1. Mean sucking rate is shown for 1-min sample periods before and after the habituation criterion was attained. Separate curves are shown for experimental (melodic contour change at habituation) and control (no change) groups. Vertical bars represent \pm one standard error.

Sound presentation was contingent upon the infant's sucking behavior. Each time a high-amplitude suck was detected by the computer, a ramped gate was opened that allowed the continuous sound sequence on tape to be heard. The ramp lasted half the duration of a four-tone cycle. After the ramp, a minimum of three complete cycles was presented. If no further sucking was detected during this time the sequence was ramped off over half a cycle. Each high-amplitude suck detected during sound presentation resulted in the continuation of sound presentation for 1.2 s following the suck. In general, sucking behavior in newborns occurs in bursts of regular sucking that last several seconds and have a rate of 1.5 to 2 sucks/s. Thus when the infants in our study maintained a rate of at least 0.83 sucks/s during a burst, the sound would be presented continuously during the burst.

B. Results and discussion

The data for three subjects (two control and one experimental) were subsequently rejected since their sucking rate had fallen below seven sucks/min in the 1-min periods either preceding or succeeding the habituation point. This rejection is justified by the fact that a very low sucking rate may give rise first to very few stimulus cycles at a crucial point in the experiment, and second to an exceedingly long silent interval between the temporally adjacent presentations of the pre- and post-habituation stimulus sequences. The average sucking rates for the two groups (18 control, 19 experimental) in the three 1-min periods preceding and the two 1-min periods succeeding the habituation point are shown in Fig. 2. No difference between groups was found in the three 1-min pe-

riods prior to habituation according to a repeated measures ANOVA [$F(2,70) < 1$], indicating similar behavior during the habituation phase in both groups. It appears that the sucking rate of the experimental group increases slightly more than that of the control group after the habituation point, due to the fact that rate for the experimental group is slightly lower in the pre-habituation period. In order to test the amount of change across the habituation point, a mixed ANOVA was performed with sucking rate as dependent variable and with within-subjects factor Period (1-min periods before and after habituation) and between-subjects factor group (control, experimental). The change in average rate across the habituation point for the control group was not significantly different from that for the experimental group [$F(1,35) < 1$]. Identical results are obtained if average sucking rate is computed in 2-min periods preceding and succeeding habituation, although additional subjects must be removed whose sucking rates do not reach criterion in the newly included 1-min periods. These results may be interpreted as indicating that newborn infants cannot discriminate these rising and falling sequences.

In the absence of data on newborns for this kind of discrimination, we chose the values of pitch interval and tempo used in this experiment since they generally give good rising/falling contour discrimination in adults. It is possible that either the sequences were too rapid or the pitch intervals too small (or both) for the infants to be able to acquire differentiable mental representations of these contours. Our goal in this experiment was not to study the respective effects of each of these sequence variables, but to find a melodic pattern for which original and retrograde versions could be discriminated in order to test our main hypothesis concerning stream organization. Therefore, a lower tempo and greater interval size were employed in experiment 2.

III. EXPERIMENT 2

A. Method

The method was identical to that in experiment 1. The stimulus configuration was identical except for doubling the inter-tone onset time to 200 ms and increasing the pitch interval between adjacent tones to five semitones. The pitch sequences used were thus E3-A3-D4-G4 and its retrograde. When the infant sucked on the pacifier, the continuously running sequence was ramped on over 400 ms (half-cycle of the sequence) and stayed on for an additional two cycles before ramping off over 400 ms if an additional suck did not occur. A rate of at least 0.5 sucks/s thus resulted in continuous sound presentation during a burst of sucking. Twenty subjects completed the experiment in each of the experimental and control groups. Data for 44 additional subjects were rejected: 17 did not habituate within 15 min, 13 produced insufficient or irregular sucking or spat out the pacifier, three fell asleep, and 11 cried or became agitated.

B. Results and discussion

The data for four control subjects were subsequently rejected since their sucking rate had fallen below seven sucks/min in the 1-min periods either preceding or succeeding the

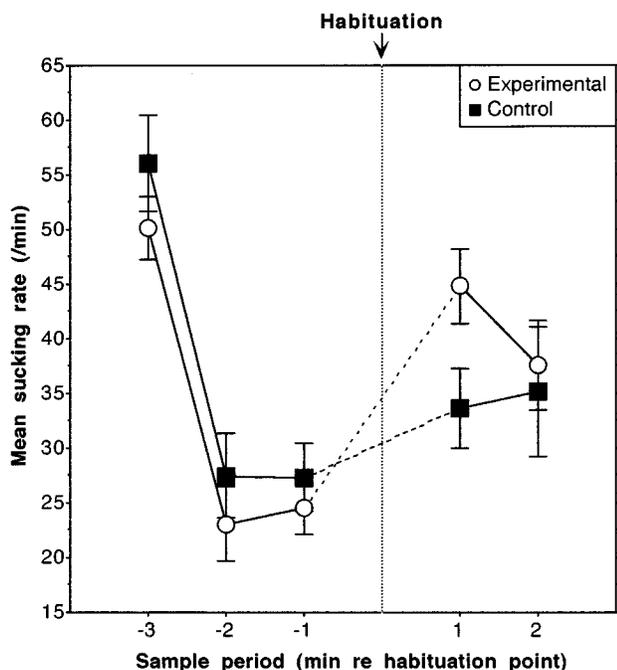


FIG. 3. Results for experiment 2 (see Fig. 2 caption).

habituation point. Data for 20 experimental and 16 control subjects were analyzed. According to a repeated measures ANOVA, no difference between groups was found in the 3-min periods prior to habituation [$F(2,68) < 1$] (Fig. 3). A Period (2) \times Group (2) mixed ANOVA was performed as in experiment 1 and revealed that the experimental group's sucking rate increased more across the habituation point than did that of the control group [$F(1,34) = 6.9, p < 0.05$]. This latter effect is slightly weaker if sucking rate is computed on 2-min periods on either side of the habituation point [$F(1,30) = 3.6, p = 0.066$]. These results indicate that newborn infants can discriminate the rising and falling melodic contours used in this experiment).

Comparisons of the data for experiments 1 and 2 by way of planned contrasts within a between-subjects ANOVA [experiment (2) \times group (2)] with the difference in sucking rate between pre-habituation and post-habituation periods as dependent variable clearly show that: (1) across the two experiments performance was similar in the control groups [$F(1,69) < 1$] and marginally different between experimental groups [$F(1,69) = 3.5, p = 0.065$, although this difference is nonsignificant if two-minute periods are used to compute mean sucking rate: $F(1,61) = 2.4$], and (2) differences between experimental and control groups were not significant for experiment 1 [$F(1,69) < 1$] but were significant for experiment 2 [$F(1,69) = 8.0, p < 0.01$]; the same pattern of results was found for sucking rates computed on 2-min periods. This suggests a greater sensitivity to melodic contour change for the slower tempo and larger pitch interval patterns. This difference is unlikely to be related to psychoacoustic limits of frequency discrimination. Frequency resolution is better than or equal to 4% for pure tones at 3 months of age when the sounds are presented at 40 dB SL (Olsho *et al.*, 1987). Frequency resolution is lower at higher fre-

quencies in early infancy and increases over the first year of life, but there is a much smaller effect of development at lower frequencies that are comparable to the fundamental frequencies in our stimuli (Spetner and Olsho, 1990). Further, Olsho (1985) has shown similar psychoacoustic tuning curves in 4-month-olds and adults, although it is unknown whether newborns also have similar tuning to adults.

The better discrimination of slower sequences may also be due to limits in temporal resolution. Gap detection studies by Werner *et al.* (1992) have measured thresholds at about 60 ms in the period of 3–12 months, while thresholds measured in adults are generally less than 10 ms. These thresholds measured in the presence of low-pass noise are up to 100 ms for 3- to 6-month-olds. Corresponding thresholds ($d' \approx 1.0$) in data from Trehub *et al.* (1995) for 6.5-month-old infants were just under 30 ms for tone pips. Again, to our knowledge no data are available on temporal resolution in newborns. The inter-tone intervals in our experiments 1 and 2 were 15 and 115 ms, respectively. Marean and Werner (1991) have shown 20 dB of forward masking of a 1-kHz pure tone by a broadband noise with an inter-tone interval of 20 ms and 12 dB of masking for an interval of 100 ms. While these data are for older infants, the resolution may be even worse at birth. Further research will be needed to tease apart the relative importance of these two factors in the perception of melodic contours by newborn infants and their applicability to discrimination of melodic sequences such as those employed in this study.

IV. EXPERIMENT 3

Having found conditions in which the infants can discriminate the rising and falling sequences, experiment 3 was then run with these same values in order to test the streaming hypothesis. We hypothesized that infants would organize the stimulus sequence into two two-note streams, each of which would not be distinguishable from its retrograde version.

A. Method

The method was identical to that in the previous two experiments. The tempo and pitches of the stimulus sequences were identical to those from experiment 2, but the 2/2 configuration of timbres and spatial positions was used (Fig. 1). Twenty-four subjects completed the experiment in each of the experimental and control groups. Data for 45 additional subjects were rejected: 10 did not habituate within 15 min, 18 produced insufficient or irregular sucking or spat out the pacifier, 5 fell asleep, 10 cried or became agitated, and 2 were removed due to experimenter error.

B. Results and discussion

The data for five subjects (four control and one experimental) were subsequently rejected since their sucking rate had fallen below seven sucks/min in the 1-min periods either preceding or succeeding the habituation point. The data for the two groups (20 control, 23 experimental) are presented in Fig. 4. No difference between groups was found in the 3-min periods prior to habituation [$F(2,82) = 1.3, n.s.$]. A mixed ANOVA [period (2) \times group (2)] reveals that the experi-

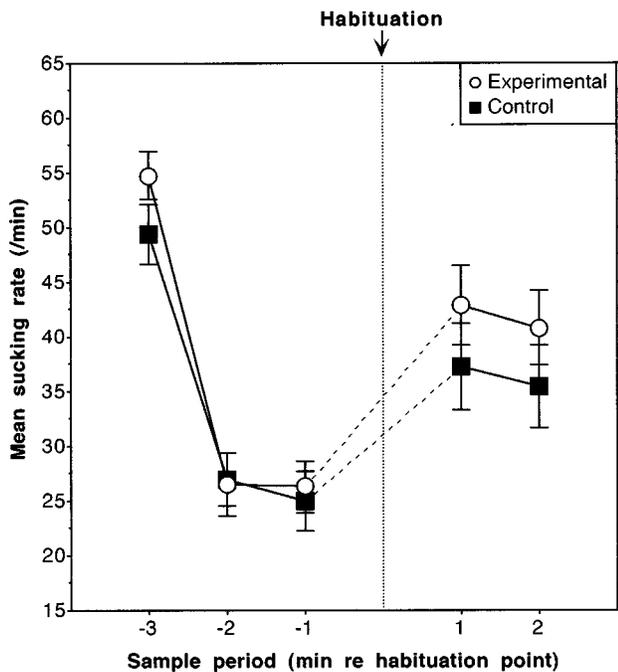


FIG. 4. Results for experiment 3 (see Fig. 2 caption).

mental group's mean sucking rate did not increase more across the habituation point than did that of the control group [$F(1,41) < 1$], indicating that they did not discriminate rising and falling contours in the 2/2 configuration. Identical results were found for sucking rates computed on 2-min periods before and after the habituation point.

Data from experiments 2 and 3 were compared by way of planned contrasts within a between-subjects ANOVA [experiment (2) \times group (2)] with the difference in sucking rate between pre-habituation and post-habituation periods as dependent variable. Although rates were similar in the control and experimental groups across the two experiments [$F(1,75) < 1$ in both cases], differences between experimental and control groups were significant for experiment 2 [$F(1,75) = 5.2, p < 0.05$] but were not significant for experiment 3 [$F(1,75) < 1$]. The significant contrast for experiment 2, derived from 1-min sucking periods, is weakened somewhat when 2-min periods are used. This difference may be due to the transient increase in response to novelty of the sequence change for the experimental group as can be seen in Fig. 3. These results suggest that newborns do not discriminate rising from falling patterns when their events are distributed in the 2/2 configuration on the basis of timbre and spatial position.

We can rule out a number of potential explanations for this difference between experiments: (1) The effect cannot be ascribed to general differences in the stimuli since the only difference between experiments 2 and 3 is that one note in the pattern changed timbre and position (transforming the 3/1 configuration into the 2/2 configuration). Otherwise there was no change in stimulus complexity: the melodic patterns had the same pitches, the same intensity, the same tempo, and each had two timbres and two spatial positions. (2) Differences in time elapsed between the last presentation of an

habituation stimulus and the first presentation of a new stimulus for the experimental group can also be ruled out. A comparison of critical inter-stimulus intervals across the habituation point show no significant differences between control and experimental groups within each experiment [experiment 2: unpaired $t(34) = -0.09$, n.s.; experiment 3: $t(41) = -0.48$, n.s.] nor differences between corresponding groups across the two experiments [control: $t(34) = 1.00$, n.s.; experimental: $t(41) = 0.95$, n.s.]. (3) Finally, differences between experiments in newborns' overall response rates are not responsible for the effect either. The planned contrasts described above demonstrate that there was no difference in global response rate between corresponding subject groups in the two experiments.

The most plausible explanation for the fact that the melodic contour is discriminated in one pattern and not the other is that sequential organization processes operating on the basis of timbre and spatial position do not give perceptual access to a discriminable contour in the 2/2 configuration, but do in the 3/1 configuration. This conclusion is weakened somewhat by the lack of significant interaction in the across-experiment ANOVA, but the results of the planned contrasts are consistent with the stream-segregation hypothesis.

V. EXPERIMENT 4

Experiment 4 was conducted to verify that adult performance on the same stimuli in an explicit discrimination task would give similar results to those obtained with the newborns.

A. Method

In the experiments with the newborns, stimulus presentation was contingent upon high-amplitude sucking. The time interval between the last presentation of the habituation sequence and the first post-habituation sequence varied from 0.3 to 45.3 s across all three experiments ($M = 11.2$ s, $s.d. = 11.7$ s). The upper limit on this interval would be constrained by our low-rate rejection criterion. In order to simulate the same kind of variation for adult listeners, we therefore decided to present inter-sequence silences (ISIs) of 5, 15, and 25 s for each experimental condition. These values are longer than 37%, 70%, and 84%, respectively, of all ISIs across the habituation point in the newborn experiments.

The experiment was conducted in two 45-min sessions, one with the fast tempo/small pitch interval condition and one with the slow tempo/large interval condition. The order of presentation of the sessions was counterbalanced across two groups of six nonmusician listeners. In each session 48 trials were presented composed of two configurations (3/1, 2/2), three inter-sequence silences (5, 15, 25 s), four comparisons (rising/rising, falling/falling, rising/falling, falling/rising), and two repetitions.

The subjects heard a warning signal followed 2 s later by the initial sequence which faded in over 2 s, played 10 cycles at full level, and then faded out over 2 s. After a variable silence, the second sequence was presented in the same manner. The subject was to judge whether the order of the pitches (i.e., the pitch contour) in the two sequences was

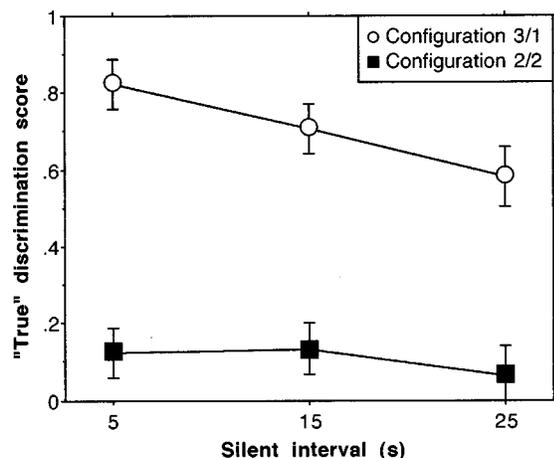


FIG. 5. Results for experiment 4. Mean “true” discrimination scores (hit rate minus false alarm rate) are shown as a function of duration of the silent interval separating the two sequences to be compared in a trial. (Chance performance is zero for this score.) The data for the two stimulus configurations (3/1 and 2/2), averaged across tempo/interval size conditions, are shown as separate curves. Vertical bars represent \pm one standard error.

the same or different. One “same” and one “different” pair were presented for each initial sequence in each experimental condition. No feedback was given concerning the correct response.

The subjects in each group were tested collectively. They were seated in a sound-treated room in front of two loudspeakers that formed an azimuthal angle of between 60° and 90° depending on the subject’s distance from the loudspeakers. The trumpet sound was always presented in the right speaker and the vibraphone in the left. Subjects marked their responses on an answer sheet.

B. Results and discussion

For each condition and subject, a “true discrimination” score was computed by subtracting the false-alarm rate from the hit rate across comparisons and repetitions. This gives a performance score that varies between 0 (chance performance) and 1 (perfect performance) unless listeners systematically respond incorrectly in which case the score can be negative. The mean scores across tempo/interval and presentation order are shown in Fig. 5 as a function of inter-sequence silence. The scores were submitted to a four-way ANOVA with repeated within-subjects factors tempo/interval(2) \times configuration(2) \times silence(3) and between-subjects factor presentation order of sessions (2). The only significant effects were silent interval [$F(2,20) = 3.94$, $p < 0.05$], indicating that performance decreases slightly overall when the delay between target and comparison sequence is long, and timbre/space configuration [$F(1,10) = 74.48$, $p < 0.0001$], demonstrating clearly that subjects perform well with the 3/1 configuration and very poorly with the 2/2 configuration (0.70 vs 0.11 globally), as was hypothesized at the beginning. Thus as measured in a paradigm requiring an explicit response from adult subjects, results similar to those obtained with newborns are obtained, sug-

gesting in both cases that the sequences are organized into two streams on the basis of timbre and spatial location of the sound events.

VI. GENERAL DISCUSSION

These results show that newborns can discriminate rising and falling melodic patterns under certain conditions. They are able to make the discrimination, as measured with the non-nutritive sucking paradigm, for large pitch interval/slow tempo sequences organized in the 3/1 configuration (experiment 2), but they do not discriminate similar patterns with small interval and fast tempo conditions (experiment 1). Adult subjects easily discriminate the 3/1 patterns in both conditions using a classic 2AFC task (experiment 4). Although different paradigms were used for adults and newborns, the differences in performance indicate that melodic pattern discrimination has some limits at birth that improve with age. The available literature has not measured frequency and temporal resolution in newborns, but data from 3-month-olds suggest that the limit in our study is probably related more to poor temporal resolution than to poor frequency resolution. Our study extends work on pattern perception in older infants by Trehub *et al.* (1987), indicating that at least some of the capacities demonstrated in their study are present at birth.

Neither newborns nor adults discriminated original from retrograded 2/2 configurations (experiments 3 and 4, respectively). This result, taken together with the capacity to discriminate similar 3/1 configurations in experiment 2, suggests that the ability to discriminate melodic sequences depends on the way the events are organized into streams on the basis of timbre and/or spatial position of the sources. Work by Hartmann and Johnson (1991) on adults using an interleaved melody recognition task, indicates that it is probably the timbre difference that is primarily responsible. Our study extends work reported by Demany (1982) and Fassbender (1993) for similar kinds of sequences presented to infants of 1.5–5.5 months of age. We are inclined to interpret the data as indicating that stream organization mechanisms are present at birth. However, the statistical weakness of the critical comparisons makes it clear that this kind of study needs to be replicated with other response paradigms and perhaps performed with a paradigm in which discrimination is required to demonstrate streaming to confirm the present results in which streaming is inferred from a lack of discrimination.

This study is the first to investigate stream organization in newborns. Methodological limitations of research with newborn babies could be partly responsible for the lack of work on such questions. Most experiments address questions of the type “Are babies able to discriminate A from B?” A and B being single stimulus events or categories. Questions about how newborns construct percepts and organize complex stimuli are far more difficult to address, as answers to them are often inferred from discrimination paradigms. These latter present difficulties for long sequences even in adults, as is witnessed by the relative lack of such research in this journal. In spite of the tenuousness in interpreting newborn data on such complex perceptual processing, the present data suggest, in accord with previous studies, that the most

basic processes for stream organization are operative very early in life. Babies can build auditory streams on the basis of frequency proximity, intensity similarity, and spectral similarity (Demany, 1982; Fassbender, 1993). The present study suggests that, in addition, complex timbral and spatial properties of sound sources may also be used as cues to form auditory streams from sequences of events and to differentiate the streams thus formed. These presumably unlearned processes (even if they are more affected by certain qualities of the stimuli such as pitch interval and tempo than corresponding processes in adults) would allow infants to perceptually structure their acoustic environment.

We know that infants perceive visual objects and events according to innate constraints that allow them to organize their knowledge of objects and events earlier than they could do on the basis of their own experience with these objects (Baillargeon *et al.*, 1990; Spelke *et al.*, 1992). There is almost no comparable data on the way infants acquire knowledge about the *acoustic* world. Some innate predisposition seems to be at work for perceiving speech (Pinker, 1994). But what about more general-purpose mechanisms? To what extent is the human auditory system innately predisposed to use different acoustic cues in such a way that auditory scenes could be analyzed in terms of objects or coherent sources and not as a collection of acoustic dimensions that the infant would have to learn to relate and to combine through experience? Our results are compatible with the hypothesis that some basic processes are in place at birth, even if not completely developed. More research is needed to better understand how these basic processes participate in the early development of more “strategic” or “heuristic” processes such as those that are characterized as “schema-based” by Bregman (1990). For example, Newman and Jusczyk (1996) recently tested 7.5-month-old infants’ abilities to extract speech information (words) delivered by a voice in the presence of another, competing voice, speaking simultaneously at different intensities. Several experiments showed that the infants could recognize target words, indicating that they were able to segregate two speech streams and to selectively and continuously attend to the target speech, at least when it was more intense than the background speech.

Given the extensive research with adults on such questions, it is imperative to explore infants’ primary abilities in auditory cognition. The present research will perhaps have had the merit of inciting further work in this little explored domain.

ACKNOWLEDGMENTS

This work was supported in part by a fellowship from the Fyssen Foundation to SM tenured at the Laboratoire de Sciences Cognitives et Psycholinguistique. We would like to thank Jacques Mehler for having suggested the idea of doing this research and for numerous illuminating discussions. Jacqueline Bobrow helped run subjects, and Laurent Demany and two anonymous reviewers made several helpful comments.

¹While we conceive of the dimension of discrimination as being “contour” in this study, it should be pointed out that we cannot separate melodic interval pattern discrimination from discrimination of the contour of ups and downs in the melody here since the two covary, i.e., we do not present similar contours that vary in interval pattern and absolute pitch as did Trehub *et al.* (1987).

- Baillargeon, R., Graber, M., DeVos, J., and Black, J. C. (1990). “Why do young infants fail to search for hidden objects?,” *Cognition* **36**, 255–284.
- Bregman, A. S. (1978). “Auditory streaming is cumulative,” *J. Exp. Psychol.* **4**, 380–387.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).
- Bregman, A. S., and Campbell, J. (1971). “Primary auditory stream segregation and perception of order in rapid sequences of tones,” *J. Exp. Psychol.* **89**, 244–249.
- Bregman, A. S., Liao, C., and Levitan, R. (1990). “Auditory grouping based on fundamental frequency and formant peak frequency,” *Can. J. Psychol.* **44**, 400–413.
- Clifton, R. K. (1992). “The development of spatial hearing in human infants,” in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (American Psychological Association, Washington, DC), pp. 135–158.
- Demany, L. (1982). “Auditory stream segregation in infancy,” *Inf. Behav. Dev.* **5**, 261–276.
- Fassbender, C. (1993). *Auditory Grouping and Segregation Processes in Infancy* (Kaste Verlag, Norderstedt).
- Ferland, M. B., and Mendelson, M. J. (1989). “Infants’ categorization of melodic contour,” *Inf. Behav. Dev.* **12**, 341–355.
- Floccia, C., Christophe, A., and Bertoncini, J. (1997). “High-amplitude sucking and newborns: The quest for underlying mechanisms,” *J. Exp. Child Psychol.* **64**, 175–198.
- Hartmann, W. M., and Johnson, D. (1991). “Stream segregation and peripheral channeling,” *Music Percept.* **9**, 155–184.
- Jusczyk, P. W. (1985). “The high-amplitude sucking technique as a methodological tool in speech perception research,” in *Measurement of Audition and Vision in the First Year of Life: A Methodological Overview*, edited by G. Gottlieb and N. A. Krasnegor (Ablex, Norwood, NJ), pp. 195–222.
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language* (MIT, Cambridge, MA).
- Marean, G. C., and Werner, L. A. (1991). “Forward masking functions of 3-month-old infants,” *J. Acoust. Soc. Am.* **89**, 1914(A).
- McAdams, S., and Bregman, A. S. (1979). “Hearing musical streams,” *Comput. Music J.* **3**, 26–43.
- McAdams, S., Bertoncini, J., and Bobrow, J. (1990). “Organization and discrimination of repeating sound sequences by newborn infants,” *J. Acoust. Soc. Am. Suppl.* **1** **88**, S91.
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (1995). “Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes,” *Psychol. Rev.* **58**, 177–192.
- Newman, R. S., and Jusczyk, P. W. (1996). “The cocktail party effect in infants,” *Percept. Psychophys.* **58**, 1145–1156.
- Olsho, L. W. (1985). “Infant auditory perception: Tonal masking,” *Inf. Behav. Dev.* **8**, 371–384.
- Olsho, L. W., Koch, E. G., and Halpin, C. F. (1987). “Level and age effects in infant frequency discrimination,” *J. Acoust. Soc. Am.* **82**, 454–464.
- Pinker, S. (1994). *The Language Instinct* (Morrow, New York).
- Spelke, E. S. (1990). “Principles of object perception,” *Cogn. Sci.* **14**, 29–56.
- Spelke, E. S., Breinlinger, K., Macomber, J., and Jacobson, K. (1992). “Origins of knowledge,” *Psychol. Rev.* **99**, 605–632.
- Spetner, N. B., and Olsho, L. W. (1990). “Auditory frequency resolution in human infancy,” *Child Dev.* **61**, 632–652.
- Trehub, S. E., and Trainor, L. J. (1993). “Listening strategies in infancy: The roots of music and language development,” in *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by S. McAdams and E. Bigand (Oxford U.P., Oxford), pp. 278–327.
- Trehub, S. E., Endman, M. W., and Thorpe, L. A. (1990). “Infants’ perception of timbre: Classification of complex tones by spectral structure,” *J. Exp. Child Psychol.* **49**, 300–313.
- Trehub, S. E., Schneider, B. A., and Henderson, J. L. (1995). “Gap detection in infants, children, and adults,” *J. Acoust. Soc. Am.* **98**, 2532–2541.

- Trehub, S. E., Thorpe, L. A., and Morrongiello, B. A. (1987). "Organization processes in infants' perception of auditory patterns," *Child Dev.* **58**, 741–749.
- van Noorden, L. P. A. S. (1977). "Minimum differences of level and frequency for perceptual fission of tone sequences ABAB," *J. Acoust. Soc. Am.* **61**, 1041–1045.
- Werner, L. A. (1992). "Interpreting developmental psychoacoustics," in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (American Psychological Association, Washington, DC), pp. 47–88.
- Werner, L. A., and Rubel, E. W. (Eds.) (1992). *Developmental Psychoacoustics* (American Psychological Association, Washington, DC).
- Werner, L. A., Marean, G. C., Halpin, C. F., Spetner, N. B., and Gillenwater, J. M. (1992). "Infant auditory temporal acuity: Gap detection," *Child Dev.* **63**, 260–272.
- Wessel, D. L., Bristow, D., and Settel, Z. (1987). "Control of phrasing and articulation in synthesis," in *Proceedings of the 1987 International Computer Music Conference* (Computer Music Association, San Francisco), pp. 108–116.