

Newborn infants can organize the auditory world

István Winkler^{*†}, Elena Kushnerenko[‡], János Horváth^{*}, Rita Čeponienė[§], Vineta Fellman^{¶||}, Minna Huotilainen[‡], Risto Näätänen[‡], and Elyse Sussman^{**}

^{*}Institute for Psychology, Hungarian Academy of Sciences, H-1394 Budapest, Hungary; [‡]Cognitive Brain Research Unit, Department of Psychology, Helsinki University, and Helsinki Brain Research Centre, FIN-00014 Helsinki, Finland; [§]Project in Cognitive and Neural Development, Language Research Center, University of California at San Diego, La Jolla, CA 92093-0013; [¶]Hospital for Children and Adolescents, Helsinki University Central Hospital, FIN-00029 Helsinki, Finland; ^{||}Department of Pediatrics, Lund University, S-22185 Lund, Sweden; and ^{**}Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461

Edited by Eugene N. Sokolov, Moscow State University, Moscow, Russia, and approved August 7, 2003 (received for review April 2, 2003)

The perceptual world of neonates is usually regarded as not yet being fully organized in terms of objects in the same way as it is for adults. Using a recently developed method based on electric brain responses, we found that, similarly to adults, newborn infants segregate concurrent streams of sound, allowing them to organize the auditory input according to the existing sound source. The segregation of concurrent sound streams is a crucial step in the path leading to the identification of objects in the environment. Its presence in newborn infants shows that the basic abilities required for the development of conceptual objects are available already at the time of birth.

We know that newborn infants can recognize their mother's voice (1), but can they distinguish their mother's voice in the presence of other sound? In general, can neonates organize sounds by their source? Segregating concurrent streams of sound is a crucial moment of sound organization (2), a prerequisite of perceiving objects and, thus, also of forming conceptual objects (3). The processes separating concurrent sound sequences use the temporal behavior of various acoustic parameters, of which spectral pitch is the most effective. In adults, fast presentation of sounds selected from two separated frequency ranges results in an unambiguous perception of two different sound streams, one for the lower- and the other for the higher-pitched sounds (2). Few studies tested whether young (although not newborn) infants can segregate concurrent streams of sound by using behavioral indicators such as the head-turning or the nonnutritive sucking response (4, 5). The results of these studies have suggested that the mechanisms of auditory-stream segregation may be present in young infants.

Our objective test of auditory-stream segregation is based on the fact that the detection of an auditory regularity often depends on how one organizes the incoming sounds. For example, when a familiar melody is interleaved with other sounds, the perception of the original tune is lost. However, if the sounds of the melody are taken from a pitch range that is distinctly different from that of the interleaved sounds, perception of the melody returns (6), because the two sets of sounds are treated as two independent sound sequences (i.e., the two sound streams have been segregated).

To determine whether newborn infants segregate two interleaved sound sequences of different frequency ranges, we measured electric brain responses elicited by infrequent deviant tones embedded in the sequence of a repetitive standard tone (the "oddball" sequence; see Fig. 1A, control condition). In this tone sequence, deviants elicit a brain response that is not present in the standard-stimulus response, the mismatch negativity (MMN) event-related brain potential (7). MMN is often followed by another deviation-related electric brain potential, the P3a component (8). These deviation-related brain responses are elicited irrespective of whether the subject performs some task with or ignores the sounds. However, the MMN response is only elicited if the frequent repetition of the standard tone has been detected by the brain. When randomly varying intervening tones are mixed together with the original oddball tone sequence, the

intervening tones prevent the brain from detecting the repetition of the standard tone (Fig. 1C, one-stream condition). Therefore it is predicted that the brain responses elicited by the standard and deviant tones will not differ from each other in this condition. Transposing the intervening tones to a frequency range that differs from that of the original tone sequence (Fig. 1E, two-stream condition) while keeping the amount of variation of these tones constant allows the brain to segregate the two sets of tones into separate sound streams. If and only if segregation takes place will the deviation-related brain responses reemerge. Therefore, the elicitation of these brain responses indicates segregation of the original and intervening tones.

This test of auditory-stream segregation, which does not require subjects to perform some task with or report their perception of the sounds, has already been tested successfully in adults (9, 10) and school-age children (11), and the results were found to be fully compatible with the subjects' perception of the sound streams (10, 11). Because deviation-related electric brain responses can be measured in newborn and older infants (12–18), auditory-stream segregation could be tested in newborns.

Methods

The study protocol was approved by the Ethics Committee of the Hospital for Children and Adolescents (infant experiments) and the Ethical Committee of the Institute for Psychology (adult experiments).

Subjects. Fourteen healthy newborn infants (2–5 days of age, gestational age 38–42 weeks, seven females) and eight young adults (18–23 years of age, six females) participated in the experiments. Written informed consent was obtained from the subjects (adult experiments) or their parents (infant experiments) after the nature and procedures of the experiment were explained to them. All subjects had normal hearing (checked with otoacoustic emissions in infants and audiometry in adults). Infants were measured in active sleep, and adults watched a self-selected movie without sound. Data from three newborn infants were rejected from analysis because of changes in the sleep stage during the recording. Another three infants' data were rejected due to an excessive number of high-amplitude electric artifacts.

Stimuli. Control condition. As depicted in Fig. 1A, simple sinusoid tones of 50-ms duration (including 5-ms linear rise and 5-ms fall times) were presented with a uniform 750-ms onset-to-onset interval. Tone frequency was 1,813 Hz. Frequent softer (61-dB sound pressure level, 90% probability) and infrequent louder

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: MMN, mismatch negativity.

[†]To whom correspondence should be addressed at: Institute for Psychology, Hungarian Academy of Sciences, P.O. Box 398, Szondi u. 83/85, H-1394 Budapest, Hungary. E-mail: winkler@cogpsyphy.hu.

© 2003 by The National Academy of Sciences of the USA

(76-dB sound pressure level, 10% probability) tones were delivered in a randomized order.

One-stream condition. As depicted in Fig. 1C, two intervening tones were introduced between consecutive tones of the control sequence, reducing the uniform tone onset-to-onset interval to 250 ms. The intervening tones equiprobably took on one of four frequency (1,655, 1,732, 1,898, or 1,986 Hz) and one of four intensity (66, 71, 81, or 86 dB sound pressure level) levels, 16 different tones altogether, and were delivered in a randomized order.

Two-stream condition. As depicted in Fig. 1E, the tone frequencies of the intervening tones were lowered to 250, 262, 287, and 300 Hz from the values used in the one-stream sequences. All other parameters were retained.

For each condition, two stimulus blocks were delivered. Each stimulus block contained 900 standard and 100 deviant tones (oddball sequence). One- and two-stream sequences contained 2,000 intervening tones in addition to the 1,000 tones of the oddball sequence. Tones were delivered through loudspeakers to infants and through headphones to adults.

Presentation-rate control experiment. This experiment, depicted in Fig. 1G, was designed to test whether shortening the onset-to-onset time from 750 ms (control condition) to 250 ms (one-stream condition) in and of itself could qualitatively modify the electric brain responses. To test this possibility, the parameters of all intervening tones were set to equal the corresponding parameters of the frequent tone (frequency, 1,813 Hz; intensity, 61 dB). All other stimulus parameters and experimental procedures were exactly as in the one- and two-stream conditions. Three stimulus blocks of this type were delivered to seven healthy newborn infants (1–4 days of age, gestational age 39–42 weeks, four females) and four stimulus blocks to eight young adults (18–24 years of age, five females). All subjects had normal hearing.

Data Collection and Measurement. The electroencephalogram was recorded from seven scalp locations (F3, F4, C3, Cz, C4, P3, and P4 of the international 10–20 system) against the common electric reference (the linked mastoids for infants and the tip of the nose for adults). The electrooculogram was recorded between electrodes placed lateral to the outer canthi and between electrodes placed above and below the right eye. Signals were digitized at 250-Hz sampling frequency (0- to 40-Hz band limits) and offline-filtered with a bandpass of 1–16 Hz. For each stimulus, epochs of 550-ms duration including a 100-ms pre-stimulus period were extracted from the continuous electroencephalogram/electrooculogram record. Epochs with a voltage change $>100 \mu\text{V}$ were rejected from further analysis. On average, 60 deviant and 370 standard responses were retained for newborns (175 deviant and 3,771 standard responses in the presentation-rate control experiment), and 180 deviant and 1,310 standard responses were retained for adults (328 deviant and 6,923 standard responses in the presentation-rate control experiment). Responses were averaged separately for different stimulus types (the standard vs. deviant tones), condition (control vs. one-stream vs. two-stream stimulus sequences), and experiment. Amplitude measurements, which were averaged from 30-ms intervals in the right frontal (F4) traces centered on the peaks identified in the group-average difference waveforms, were referred to the mean voltage in the -20 - to $+20$ -ms period serving as the baseline.

Results

Control Condition. In infants, the deviant tones elicited an electric brain response that was positively displaced in the 180- to 400-ms poststimulus latency range, compared with the response to the standard tone (Fig. 1B Left). In adults, the deviant tones elicited a response that was negatively displaced in the 100- to 180-ms

(MMN) and positively in the 180- to 280-ms (P3a) latency range, compared with the response elicited by the standard tone (Fig. 1B Right). The differences between the responses elicited by the deviant and standard tones indicate that the deviant tones were processed differently from the standard tones in both groups of subjects. The actual brain responses of newborns and adults, of course, are quite different (for reviews of the maturation of electric brain responses, see refs. 19–21) and, because newborns were asleep during the experiment, only a positive electric brain-response difference was elicited in them (14), whereas adults showed both a negative (MMN) and a positive (P3a) response difference.

One-Stream Condition. In newborns as well as in adults, the responses elicited by the deviant and standard tones did not differ significantly from each other (Fig. 1D). It seems that the loudness variation introduced by the intervening tones eliminated the differential processing of the standard and deviant tones that only differed in this feature. The lack of differential processing demonstrates that all tones were integrated into a single sound stream, which allowed the intervening tones to interfere with the detection of the repetition of the standard tone.

Two-Stream Condition. In both newborn infants and adults, the electric brain responses elicited by the standard and deviant tones (Fig. 1F) were very similar to the respective responses measured in the control condition (Fig. 1B). Thus, in the two-stream condition, again, the deviant tones were processed differently from the standard tones, just as it was in the absence of the intervening tones (in the control condition). This result shows that the control and intervening tones were segregated into separate sound streams, which eliminated the interfering effect of the intervening tones.

It should be noted that the amount of variation of the intervening tones was kept constant across the one- and two-stream conditions. Intensity varied exactly the same way in the two conditions, whereas the frequency variation was proportionally equal between the conditions. Moreover, variation in frequency would not have affected the elicitation of different brain responses by the standard and deviant tones, because previous studies showed that variation in one feature does not interfere with the elicitation of change-related brain potentials by deviations in another feature (22, 23). Therefore, the different pattern of results obtained in the one- and two-stream conditions was caused by the different amounts of frequency separation between the standard/deviant and intervening tones. The intervening tones could only interfere with the detection of the repetition of the standard tones when they were grouped together with them but not when they were segregated into a separate sound stream (2, 6).

Presentation-Rate Control Experiment. Deviant tones elicited discriminative electric brain responses in both groups of participants (Fig. 1H). The morphology of the event-related potential responses was quite similar to those recorded in the control and two-stream conditions of the main experiment in both newborn babies and adults. This result demonstrates that the faster presentation rate alone could not have caused the different pattern of results observed in the one-stream and two-stream conditions.

The observed differences across the three conditions of the main experiment were significant in both groups of subjects. ANOVA tests revealed a significant effect of condition on the voltage difference between the responses to deviant and standard tones [in newborns, $F(2,14) = 4.86$, $P < 0.05$, Greenhouse–Geisser $\epsilon = 0.74$ for the mean voltage difference in the 288- to 318-ms poststimulus interval; in adults, $F(2,14) = 25.26$, $P <$

0.001, $\epsilon = 0.58$ for the early, 130- to 160-ms interval and $F(2,14) = 4.38$, $P < 0.05$, $\epsilon = 0.76$ for the late, 215- to 245-ms interval]. Post hoc Newman-Keuls tests revealed significant differences between the responses elicited in the one-stream and the other two conditions ($P < 0.05$ for both comparisons in newborns and $P < 0.005$ for both comparisons in adults, early interval). In the adults but not in the newborns, the voltage difference measured in the control condition was significantly more negative in the early interval than that in the two-stream condition ($P < 0.005$).

Discussion

The present results demonstrate that auditory streaming occurs in newborn infants. By using the same paradigm as in the current experiment, our previous studies in adults (10) and school-age children (11) showed full correspondence between the elicitation of deviance-related brain responses and the perception of separate auditory streams for the oddball sequence and the intervening tones. Furthermore, the difference found in the electric brain responses could not have been caused by frequency separation between the control (standard and deviant) and intervening tones *per se*. With a constant frequency separation between high and low tones, the presentation rate controls whether the high and low tones are perceived as two separate sound streams (fast presentation rates) or as a single stream that includes all tones (slow presentation rates) (2). In full accordance with this defining characteristic of auditory streaming, it has been shown that, when sequences of high and low tones with separate regular characteristics were mixed together, increasing the rate of stimulus presentation (while frequency separation remained constant) resulted in the elicitation of deviance-related brain responses by occasional violations of one or the other acoustic regularity (9). Therefore, we conclude that the current results demonstrated the operation of auditory-stream segregation mechanisms in newborn infants.

The MMN brain response is elicited when a sound mismatches the neural representation encoding the regularities of the preceding sound sequence. In the control condition of the current experiment, this representation encoded all of the features of the repetitive standard tone, including its intensity level. The louder

deviant tones mismatched this representation and, thus, elicited the deviation-related brain responses. However, loudness was not a regular feature of the one-stream condition, because intensity varied randomly in the composite tone sequence (the oddball sequence plus the intervening tones). Therefore, in this condition, the deviant tones did not mismatch the neural representations of auditory regularities, because these representations did not encode intensity and the deviant tone was identical to the standard tone in every other feature. In the two-stream condition, in which the high-pitched tones of the oddball sequence were segregated from the low-pitched intervening tones, two separate representations were formed, one for each of the two auditory streams (24). The high stream was, of course, identical to the control-condition tone sequence. Therefore, the regularity representation formed for this stream included the intensity of the standard tone, causing deviant tones to elicit the deviation-related brain responses.

The operation of auditory stream segregation in newborn infants signifies that neonates possess the perceptual mechanisms necessary to separate the sound sources in their environment such as distinguishing their mother's voice from other concurrent sounds. Although the segregation of voice streams is based as much on timbre as on pitch cues, the operation of stream-segregation processes based on pitch separation suggests that the basic brain mechanisms responsible for representing multiple simultaneously active sound sources must already be functional at the time of birth. This ability underlies the infant's orientation in the world and provides the basis for selecting coherent subsets from the wealth of incoming information, thus enabling the development of cognitive abilities such as selective attention, speech perception (distinguishing speech from non-speech sounds and separating concurrent streams of speech from each other), social skills, and memory (by distinguishing and, subsequently, correctly representing objects) (25, 26).

We thank Drs. Gergely Csibra, György Gergely, and Judit Gervai for helpful comments on early versions of the manuscript. This research was supported by Hungarian National Research Fund Grant OTKA T034112; Academy of Finland Grants 77322, 79405, and 80820; and National Institutes of Health Grant R01 DC04263.

- DeCasper, A. J. & Fifer, W. P. (1980) *Science* **208**, 1174–1176.
- Bregman, A. S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).
- Leslie, A. M., Xu, F., Tremoulet, P. D. & Scholl, B. J. (1998) *Trends Cognit. Sci.* **2**, 10–18.
- Demany, L. (1982) *Infant Behav. Dev.* **5**, 261–276.
- McAdams, S. & Bertoncini, J. (1997) *J. Acoust. Soc. Am.* **102**, 2945–2953.
- Dowling, W. J. (1973) *Percept. Psychophys.* **14**, 37–40.
- Näätänen, R. (1990) *Behav. Brain Sci.* **13**, 201–288.
- Friedman, D., Cycowicz, Y. M. & Gaeta, H. (2001) *Neurosci. Biobehav. Rev.* **25**, 355–373.
- Sussman, E., Ritter, W. & Vaughan, H. G., Jr. (1999) *Psychophysiology* **36**, 22–34.
- Winkler, I., Horváth, J., Teder-Sälejärvi, W. A., Näätänen, R. & Sussman, E. (2003) *Cognit. Affect. Behav. Neurosci.* **3**, 57–77.
- Sussman, E., Čeponienė, R., Shestakova, A., Näätänen, R. & Winkler, I. (2001) *Hear. Res.* **153**, 108–114.
- Alho, K., Saino, K., Sajaniemi, N., Reinikainen, K. & Näätänen, R. (1990) *Electroencephalogr. Clin. Neurophysiol.* **77**, 151–155.
- Dehaene-Lambertz, G. & Dehaene, S. (1994) *Nature* **28**, 293–294.
- Friederici, A. D., Friedrich, M. & Weber, C. (2002) *NeuroReport* **13**, 1251–1254.
- Kushnerenko, E., Čeponienė, R., Balan, P., Fellman, V. & Näätänen, R. (2002) *NeuroReport* **13**, 1843–1848.
- Leppänen, P. H. T., Eklund, K. M. & Lyytinen, H. (1997) *Dev. Neuropsychol.* **13**, 175–204.
- Morr, M. L., Shafer, V. L., Kreuzer, J. A. & Kurtzberg, D. (2002) *Ear Hear.* **23**, 118–136.
- Trainor, L. J., Samuel, S. S., Desjardins, R. N. & Sonnadara, R. R. (2001) *NeuroReport* **12**, 2443–2448.
- Kurtzberg, D., Vaughan, H. G., Jr., Courchesne, E., Friedman, D., Harter, M. R. & Putnam, L. E. (1984) *Ann. N.Y. Acad. Sci.* **425**, 300–318.
- Kushnerenko, E., Čeponienė, R., Balan, P., Fellman, V., Huottilainen, M. & Näätänen, R. (2002) *NeuroReport* **13**, 47–51.
- Vaughan, H. G., Jr., & Kurtzberg, D. (1991) in *Minnesota Symposia on Child Psychology*, eds. Gunnar, M. R. & Nelson, C. A. (Erlbaum, Hillsdale, NJ), Vol. 24, pp. 1–36.
- Gomes, H., Ritter, W. & Vaughan, H. G., Jr. (1995) *J. Cognit. Neurosci.* **7**, 81–94.
- Winkler, I., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M. & Näätänen, R. (1990) *Psychophysiology* **27**, 228–235.
- Ritter, W., Sussman, E. & Molholm, S. (2000) *NeuroReport* **11**, 61–63.
- DeCasper, A. J., Lecanuet, J. P., Busnel, M. C., Granierdeferre, C. & Maugeais, R. (1994) *Infant Behav. Dev.* **17**, 159–164.
- Xu, F. & Carey, S. (1996) *Cognit. Psychol.* **30**, 111–153.