



Newborn infants detect the beat in music

István Winkler^{a,b,1}, Gábor P. Háden^{a,c}, Olivia Ladinig^d, István Sziller^e, and Henkjan Honing^d

^aInstitute for Psychology, Hungarian Academy of Sciences, H-1394 Budapest, P.O. Box 398, Hungary; ^bInstitute of Psychology, University of Szeged, H-6722 Szeged, Petőfi S. sgt. 30-34, Hungary; ^cDepartment of Cognitive Science, Budapest University of Technology and Economics, H-1111 Budapest, Stoczek u. 2, Hungary; ^dMusic Cognition Group, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Postbus 19268, 1000 GG Amsterdam; and ^eFirst Department of Obstetrics and Gynaecology, Semmelweis University, H-1088 Budapest, Baross u. 27, Hungary

Edited by Dale Purves, Duke University Medical Center, Durham, NC, and approved December 23, 2008 (received for review September 10, 2008)

To shed light on how humans can learn to understand music, we need to discover what the perceptual capabilities with which infants are born. Beat induction, the detection of a regular pulse in an auditory signal, is considered a fundamental human trait that, arguably, played a decisive role in the origin of music. Theorists are divided on the issue whether this ability is innate or learned. We show that newborn infants develop expectation for the onset of rhythmic cycles (the downbeat), even when it is not marked by stress or other distinguishing spectral features. Omitting the downbeat elicits brain activity associated with violating sensory expectations. Thus, our results strongly support the view that beat perception is innate.

event-related brain potentials (ERP) | neonates | rhythm

Music is present in some form in all human cultures. Sensitivity to various elements of music appears quite early on in infancy (1, 2–4), with understanding and appreciation of music emerging later through interaction between developing perceptual capabilities and cultural influence. Whereas there is already some information regarding spectral processing abilities of newborn infants (5, 6), little is known about how they process rhythm. The ability to sense beat (a regular pulse in an auditory signal; termed “tactus” in music theory; 7, 8) helps individuals to synchronize their movements with each other, such as necessary for dancing or producing music together. Although beat induction would be very difficult to assess in newborns using behavioral techniques, it is possible to measure electrical brain responses to sounds (auditory event related brain potentials, ERP), even in sleeping babies. In adults, infrequently violating some regular feature of a sound sequence evokes a discriminative brain response termed the mismatch negativity (MMN) (9, 10). Similar responses are elicited in newborns (11) by changes in primary sound features (e.g., the pitch of a repeating tone) and by violations of higher-order properties of the sequence, such as the direction of pitch change within tone pairs (ascending or descending) that are varying in the starting pitch (12). Newborns may even form crude sound categories while listening to a sound sequence (13): an additional discriminative ERP response is elicited when a harmonic tone is occasionally presented among noise segments or vice versa, suggesting a distinction between harmonic and complex sounds.

Neonates are also sensitive to temporal stimulus parameters [e.g., sound duration (14)] and to the higher-order temporal structure of a sound sequence [such as detecting periodical repetition of a sound pattern (15)]. Because the MMN is elicited by deviations from expectations (16), it is especially appropriate for testing beat induction. One of the most salient perceptual effects of beat induction is a strong expectation of an event at the first position of a musical unit, i.e., the “downbeat” (17). Therefore, occasionally omitting the downbeat in a sound sequence composed predominantly of strictly metrical (regular or “nonsyncopated”) variants of the same rhythm should elicit discriminative ERP responses if the infants extracted the beat of the sequence.

Results and Discussion of the Neonate Experiment

We presented 14 healthy sleeping neonates with sound sequences based on a typical 2-measure rock drum accompaniment pattern (S1) composed of snare, bass and hi-hat spanning 8 equally spaced (isochronous) positions (Fig. 1 *A* and *B*). Four further variants of the S1 pattern (S2–S4 and D) (Fig. 1 *C–F*) were created by omitting sounds in different positions. The omissions in S2, S3, and S4 do not break the rhythm when presented in random sequences of S1–S4 linked together, because the omitted sounds are at the lowest level of the metrical hierarchy of this rhythm (Fig. 1*A*) and, therefore, perceptually less salient (7). The 4 strictly metrical sound patterns (S1–S4; standard) made up the majority of the patterns in the sequences. Occasionally, the D pattern (Fig. 1*F*, deviant) was delivered in which the downbeat was omitted. Adults perceive the D pattern within the context of a sequence composed of S1–S4 as if the rhythm was broken, stumbled, or became strongly syncopated for a moment (18) ([Sound File S1](#)). A control sequence repeating the D pattern 100% of the time was also delivered (“deviant-control”).

Fig. 2 shows that the electrical brain responses elicited by the standard (only S2–S4; see *Methods*) and deviant-control patterns are very similar to each other, whereas the deviant stimulus response obtained in the main test sequence differs from them. The deviant minus deviant-control difference waveform has 2 negative waves peaking at 200 and 316 ms followed by a positive wave peaking at 428 ms. The difference between the deviant and the other 2 responses was significant in 40-ms-long latency ranges centered on the early negative and the late positive difference peaks (see Table 1 for the mean amplitudes) as shown by dependent measures ANOVAs with the factors of Stimulus (Standard vs. Deviant control vs. Deviant) × Electrode (C3 vs. Cz vs. C4). The Stimulus factor had a significant effect on both peaks (for the early negative waveform: $F[2,26] = 3.77, P < 0.05$, with the Greenhouse-Geisser correction factor $\epsilon = 0.85$ and the effect size $\eta^2 = 0.22$; for the positive waveform: $F[2,26] = 8.26, P < 0.01, \epsilon = 0.97, \eta^2 = 0.39$). No other main effects or interactions reached significance. Posthoc Tukey HSD pairwise comparisons showed significant differences between the deviant and the deviant-control responses in both latency ranges (with $df = 26, P < 0.05$ and 0.01 for the early negative and the late positive waveforms, respectively) and for the positive waveform, between the deviant and the standard response ($df = 26, P < 0.01$). No significant differences were found between the standard and the deviant control responses.

Author contributions: I.W., G.P.H., O.L., and H.H. designed research; G.P.H., O.L., and I.S. performed research; G.P.H. analyzed data; and I.W., I.S., and H.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: iwinkler@cogpsyphy.hu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0809035106/DCSupplemental.

© 2009 by The National Academy of Sciences of the USA

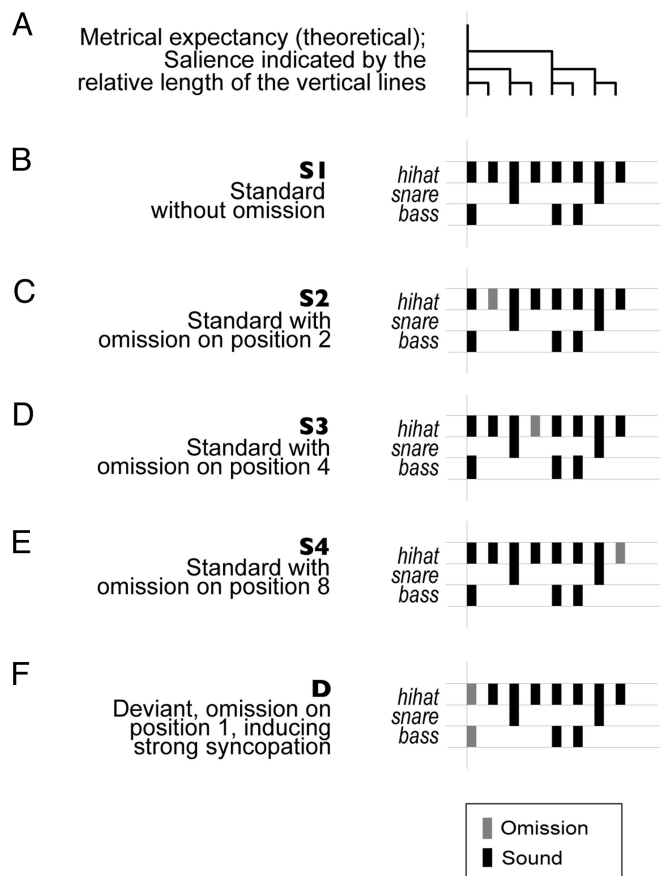


Fig. 1. Schematic diagram of the rhythmic stimulus patterns.

Results showed that newborn infants detected occasional omissions at the 1st (downbeat) position of the rhythmic pattern, but, whereas the S2–S4 patterns omitted only a single sound (the hi-hat), the D pattern omitted 2 sounds (hi-hat and bass). The double omission could have been more salient than the single sound omissions, thus eliciting a response irrespective of beat induction. However, the omission in D could only be identified as a double omission if the neonate auditory system expected both the bass and the hi-hat sound at the given moment of time. Because bass and hi-hat cooccur at 3 points in the base pattern (see Fig. 1B), knowing when they should be encountered to-

gether requires the formation of a sufficiently detailed representation of the whole base pattern in the neonate brain. In contrast, beat detection requires only that the length of the full cycle and its onset are represented in the brain. It is possible that neonates form a detailed representation of the base pattern. This would allow them not only to sense the beat, but also to build a hierarchically ordered representation of the rhythm (meter induction), as was found for adults (18). This exciting possibility is an issue for further research.

Another alternative interpretation of the results suggests that newborn infants track the probabilities of the succession of sound events (e.g., the probability that the hi-hat and bass sound event is followed by a hi-hat sound alone). However, in this case, some of the standard patterns (e.g., S2) should also elicit a discriminative response, because the omission has a low conditional probability (e.g., the probability that the hi-hat and bass sound event is followed by an omission, as it occurs in S2, is 0.078 within the whole sequence).

Finally, it is also possible that newborn infants segregated the sounds delivered by the 3 instruments, creating separate expectations for each of them. This explanation receives support from our previous results showing that newborn infants segregate tones of widely differing pitches into separate sound streams (6). If this was the case, omission of the bass sound could have resulted in the observed ERP differences without beat being induced. To test this alternative, we presented the test and the control sequences of the neonate experiment to adults, silencing the hi-hat and snare sounds. All stimulation parameters, including the timing of the bass sounds and the probability of omissions (separately for the test and the control sequences) were identical to the neonate experiment.

Results and Discussion of the Adult Control Experiment

Fig. 3 shows that the ERP responses elicited by the deviant and the control patterns are highly similar to each other. Taking the peaks where the central (C3, Cz, and C4 electrodes) deviant-minus-control difference was largest (132 and 296 ms) within the latency range in which discriminative ERP responses are found in adults, we conducted ANOVAs of similar structure as was done for the neonate measurements [dependent measures factors of Stimulus (Deviant control vs. Deviant) \times Electrode (C3 vs. Cz vs. C4); standard patterns could not be used, because they contained no omission in the bass sequence]. We found no significant main effect of Stimulus or interaction between the Stimulus and the Electrode factor ($P > 0.2$ for all tests). The only significant effect found was that of Electrode for the later latency range ($F[2,24] = 7.30, P < 0.01, \epsilon = 0.75, \eta^2 = 0.38$) However,

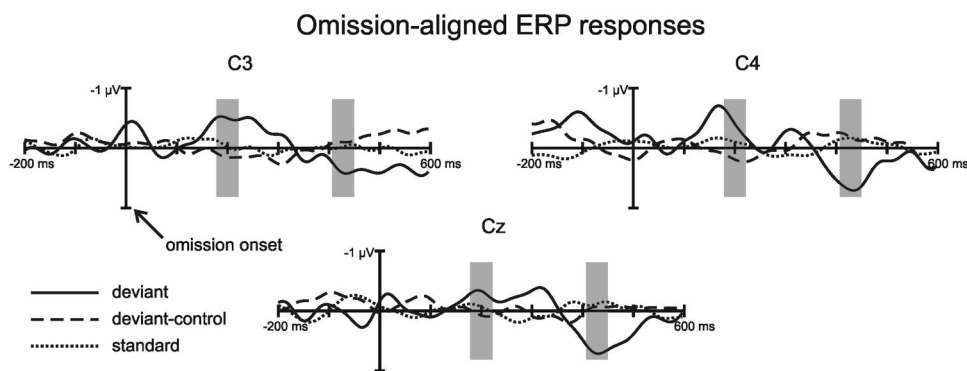


Fig. 2. Group averaged ($n = 14$) electrical brain responses elicited by rhythmic sound patterns in neonates. Responses to standard (average of S2, S3, and S4; dotted line), deviant (D; solid line), and deviant-control patterns (D patterns appearing in the repetitive control stimulus block; dashed line) are aligned at the onset of the omitted sound (compared with the full pattern: S1) and shown from 200 ms before to 600 ms after the omission. Gray-shaded areas mark the time ranges with significant differences between the deviant and the other ERP responses.

Table 1. Group averaged ($n = 14$) mean ERP amplitudes

Group	Amplitude, μV					
	180–220 ms interval			408–448 ms interval		
	C3	Cz	C4	C3	Cz	C4
Deviant	-0.50 (0.19)	-0.30 (0.22)	-0.41 (0.27)	0.38 (0.17)	0.67 (0.13)	0.67 (0.27)
Deviant-control	0.14 (0.13)	0.06 (0.16)	0.18 (0.25)	-0.10 (0.13)	-0.06 (0.16)	-0.18 (0.15)
Standard	-0.03 (0.09)	-0.06 (0.12)	-0.11 (0.09)	-0.02 (0.09)	-0.12 (0.12)	-0.16 (0.13)

SEM values are shown in parentheses.

because this effect does not include the Stimulus factor, it is not the sign of a response distinguishing the deviant from the control response.

Thus, in adults, omission of the position-1 bass sound does not result in the elicitation of discriminative ERP responses in the absence of the rhythmic context. This result is compatible with those of previous studies showing that stimulus omissions (without a rhythmic structure) only elicit deviance-related responses at very fast presentation rates (<170-ms onset-to-onset intervals; see ref. 19). In our stimulus sequences, the omitted bass sound was separated by longer intervals from its neighbors. It should be noted that adult participants elicited the MMN discriminative ERP response, when they received the full stimulus sequence (all 3 instruments) as presented to newborn babies in the neonate experiment (18).

Discussion

These results demonstrate that violating the beat of a rhythmic sound sequence is detected by the brain of newborn infants. In support of this conclusion we showed that the sound pattern with omission at the downbeat position elicited discriminative electrical brain responses when it was delivered infrequently within the context of a strictly metrical rhythmic sequence. These responses were not elicited by the D pattern per se: When the D pattern was delivered in a repetitive sequence of its own, the brain response to it did not differ from that elicited by the standards. Neither were discriminative responses simply the result of detecting omissions in the rhythmic pattern. Omissions occurring in non-salient positions elicited no discriminative responses (see the response to the standards in Fig. 2). Furthermore, the discriminative ERP response elicited by the D pattern was not caused by separate representations formed for the 3 instruments: only omissions of the downbeat within the rhythmic context elicit this response.

So it appears that the capability of detecting beat in rhythmic sound sequences is already functional at birth. Several authors consider beat perception to be acquired during the first year of life (2–4), suggesting that being rocked to music by their parents is the most important factor. At the age of 7 months, infants have been shown to discriminate different rhythms (2, 3). These results were attributed to sensitivity to rhythmic variability, rather than to perceptual judgments making use of induced beat. Our results show that although learning by movement is probably important, the newborn auditory system is apparently sensitive to periodicities and develops expectations about when a new cycle should start (i.e., when the downbeat should occur). Therefore, although auditory perceptual learning starts already in the womb (20, 21), our results are fully compatible with the notion that the perception of beat is innate. In the current experiment, the beat was extracted from a sequence comprised of 4 different variants of the same rhythmic structure. This shows that newborns detect regular features in the acoustic environment despite variance (12) and they possess both spectral and temporal processing prerequisites of music perception.

Many questions arise as a result of this work. Does neonate sensitivity to important musical features mean that music carries some evolutionary advantage? If so, are the processing algorithms necessary for music perception part of our genetic heritage? One should note that the auditory processing capabilities found in newborn babies are also useful in auditory communication. The ability to extract melodic contours at different levels of absolute pitch is necessary to process prosody. Sensing higher-order periodicities of sound sequences is similarly needed for adapting to different speech rhythms e.g., finding the right time to reply or interject in a conversation (22). Temporal coordination is essential for effective communication. When it breaks down, understanding and cooperation between partners is seriously hampered. Therefore, even if beat induction

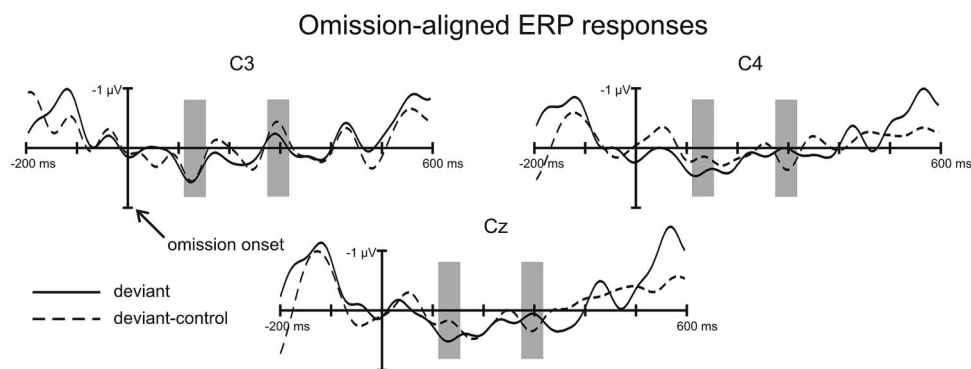


Fig. 3. Group averaged ($n = 13$) electrical brain responses elicited by the bass sound patterns in adults. Responses to deviant (D; solid line), and deviant-control patterns (D patterns appearing in the repetitive control stimulus block; dashed line) are aligned at the onset of the omitted bass sound (compared with the standard patterns: S1–S4) and shown from 200 ms before to 600 ms after the omission. Gray-shaded areas mark the time ranges in which amplitudes were measured.

is an innate capability, the origin and evolutionary role of music remains an issue for further research.

Methods

Neonate Experiment. Sound sequences were delivered to 14 healthy full term newborn infants of 37–40 weeks gestational age (3 female, birth weight 2650–3960 g, APGAR score 9/10) on day 2 or 3 post partum while the electroencephalogram was recorded from scalp electrodes. The study was approved by the Ethics Committee of the Semmelweis University, Budapest, Hungary. Informed consent was obtained from one or both parents. The mother of the infant was present during the recording.

The experimental session included 5 test sequences, each comprising 276 standard (S1–S4; Fig. 1) and 30 deviant (D) patterns and a control sequence in which the D pattern (termed deviant-control) was repeated 306 times. The control stimulus block was presented at a randomly chosen position among the test stimulus blocks. In the test sequences, S1–S4 appeared with equal probability, 22.5%, each, with the D pattern making up the remaining 10%. This is a prerequisite of the deviance detection method, which requires deviations to be infrequent within the sequence (9). The order of the 5 patterns was pseudorandomized, enforcing at least 3 standard patterns between successive D patterns. The onset-to-onset interval between successive sounds was 150 ms with 75-ms onset-to-offset interval (75-ms sound duration). Patterns in the sequence were delivered without breaks. Loudness of the sounds was normalized so that all stimuli (including the downbeat) had the same loudness.

EEG was recorded with Ag-AgCl electrodes at locations C3, Cz, and C4 of the international 10–20 system with the common reference electrode attached to the tip of the nose. Signals were off-line filtered between 1 and 16 Hz. Epochs starting 600 ms before and ending 600 ms after the time of the omission in the sound patterns (compared with the S1 pattern) were extracted from the continuous EEG record. Epochs with the highest voltage change outside the 0.1–100 μ V range on any EEG channel or on the electrooculogram (measured between electrodes placed below the left and above right eye) were discarded from the analysis. Epochs were baseline corrected by the average voltage during the entire epoch and averaged across different sleep stages, whose distribution did not differ

between the test and the control stimulus blocks. Responses to the S2–S4 patterns were averaged together, aligned at the point of omission (termed “Standard”). Responses were averaged for the D pattern separately for the ones recorded in the main test and those in the control sequences (Deviant and Deviant-control responses, respectively). For assessing the elicitation of differential ERP responses, peaks observed on the group-average difference waveforms between the deviant and deviant-control responses were selected. For robust measurements, 40-ms-long windows were centered on the selected peaks. Amplitude measurements were submitted to ANOVA tests (see the structure in *Results and Discussion of the Neonate Experiment*). Greenhouse-Geisser correction was applied. The correction factor and the effect size (partial eta-square) is reported. Tukey HSD pairwise posthoc comparisons were used. All significant results are discussed.

Adult Experiment. Fourteen healthy young adults (7 female, 18–26 years of age, mean: 21.07) participated in the experiment for modest financial compensation. Participants gave informed consent after the procedures and aims of the experiments were explained to them. The study was approved by the Ethical Committee of the Institute for Psychology, Hungarian Academy of Sciences. All participants had frequency thresholds not >20 dB SPL in the 250–4000 Hz range and no threshold difference exceeding 10 dB between the 2 ears (assessed with a Mediroll, SA-5 audiometer). Participants watched a silenced subtitled movie during the EEG recordings. One participant’s data were rejected from the analyses due to excessive electrical artifacts.

All parameters of the stimulation, EEG recording, and data analysis were identical to the neonate study except that hi-hat and snare sounds were removed from the stimulus patterns without changing the timing of the remaining sounds.

ACKNOWLEDGMENTS. We thank János Rigó for supporting the study of neonates at the First Department of Obstetrics and Gynecology, Semmelweis University, Budapest and Erika Józsa Váradiné and Zsuzsanna D’Albini for collecting the neonate and adult ERP data, respectively and to Martin Coath and Glenn Schellenberg for their comments on earlier versions of the manuscript. This work was supported by European Commission 6th Framework Program for “Information Society Technologies” (project title: EmCAP Emergent Cognition through Active Perception) Contract 013123.

- Trehub S (2003) The developmental origins of musicality. *Nat Neurosci* 6:669–673.
- Hannon EE, Trehub S (2005) Metrical categories in infancy and adulthood. *Psychol Sci* 16:48–55.
- Phillips-Silver J, Trainor LJ (2005) Feeling the beat: Movement influences infant rhythm perception. *Science* 308:1430.
- Patel A (2008) *Music, Language and the Brain* (Oxford Univ. Press, Oxford).
- Novitski N, Huotilainen M, Tervaniemi M, Näätänen R, Fellman V (2007) Neonatal frequency discrimination in 250–4000 Hz frequency range: Electrophysiological evidence. *Clin Neurophysiol* 118:412–419.
- Winkler I, et al. (2003) Newborn infants can organize the auditory world. *Proc Natl Acad Sci USA* 100:1182–1185.
- Lerdahl F, Jackendoff R (1983) *A Generative Theory of Tonal Music* (MIT press, Cambridge, MA).
- Honing H (2002) Structure and interpretation of rhythm and timing. *Dutch J Music Theory* 7:227–232.
- Kujala T, Tervaniemi M, Schröger E (2007) The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biol Psychol* 74:1–19.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) “Primitive intelligence” in the auditory cortex. *Trends Neurosci* 24:283–288.
- Alho K, Saino K, Sajaniemi N, Reinikainen K, Näätänen R (1990) Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalogr Clin Neurophysiol* 77:151–155.
- Carral V, et al. (2005) A kind of auditory “primitive intelligence” already present at birth. *Eur J Neurosci* 21:3201–3204.
- Kushnerenko E, Winkler I, Horváth J, Näätänen R, Pavlov I, Fellman V, Huotilainen M (2007) Processing acoustic change and novelty in newborn infants. *Eur J Neurosci* 26:265–274.
- Kushnerenko E, Čeponienė R, Fellman V, Huotilainen M, Winkler I (2001) Event-related potential correlates of sound duration: Similar pattern from birth to adulthood. *NeuroReport* 12:3777–3781.
- Stefanics G, et al. (2007) Auditory temporal grouping in newborn infants. *Psychophysiol* 44:697–702.
- Winkler I Interpreting the mismatch negativity (MMN) (2007) *J Psychophysiol* 21:147–163.
- Fitch WT, Rosenfeld AJ (2007) Perception and production of syncopated rhythms. *Music Percept* 25:43–58.
- Ladinig O, Honing H, Háden G, Winkler I (2009). Probing attentive and pre-attentive emergent meter in adult listeners without extensive music training. *Music Percept*, in press.
- Yabe H, et al. (1998) The temporal window of integration of auditory information in the human brain. *Psychophysiol* 35:615–619.
- Visser GHA, Mulder EIJ, Precht HFR (1992) Studies on developmental neurology in the human fetus. *Dev Pharmacol Ther* 18:175–183.
- Huotilainen M, et al. (2005) Short-term memory functions of the human fetus recorded with magnetoencephalography. *NeuroReport* 16:81–84.
- Jaffe J, Beebe B (2001) *Rhythms of Dialogue in Infancy*, Monographs for the Society for Research and Child Development (Blackwell, Boston), Vol 66, No 2.