Abstract—The ability of human fetuses to recognize their own mother’s voice was examined. Sixty term fetuses were assigned to one of two conditions during which they were exposed to a tape recording of their mother or a female stranger reading a passage. Voice stimuli were delivered through a loudspeaker held approximately 10 cm above the maternal abdomen and played at an average of 95 dB SPL. Each condition consisted of three 2-min periods: no stimulus, voice (mother or stranger), and no stimulus. Fetal heart rate increased in response to the mother’s voice and decreased in response to the stranger’s; both responses were sustained for 4 min. The finding of differential behavior in response to a familiar versus a novel voice provides evidence that experience influences fetal voice processing. It supports an epigenetic model of speech perception, presuming an interaction between genetic expression of neural development and species-specific experience.

Scientists have been fascinated with the perceptual abilities of the fetus at least since the late 1800s (see review by Kisilevsky & Low, 1998). However, systematic, sustained investigations began only in the early 1980s following the development and ready availability of sophisticated ultrasound and fetal physiological monitoring technology. It is now known that from about 30 weeks gestational age (GA), fetuses begin to react to auditory stimulation. They show reliable cardiac and body movement responses to a brief 2- to 3-s, high-pass-filtered white noise at 110 dB SPL (estimated uterine attenuation varies with increasing frequencies from 0 to about 35 dB; e.g., Querleu, Renard, Versyp, Paris-Delruce, & Crepin, 1988) during high-heart-rate-variability behavioral states (2F–4F; Kisilevsky, Pang, & Hains, 2000). By term (37 to 42 weeks GA), stimulus intensity (Kisilevsky, Muir, & Low, 1989) and frequency (Lecanuet, Granier-Defere, & Busnel, 1988) modulate the threshold and magnitude of the response.

Fetuses also show the ability to perceive and discriminate segmental speech sounds. For example, Groome et al. (1999) reported that by 36 to 40 weeks GA, fetuses respond to lower intensity, 83- to 95-dB SPL speech stimuli by displaying heart rate decelerations to vowel sounds /i/ and /u/. Near-term fetuses can also discriminate the reversal of pairs of consonant-vowel sounds, /bab/ to /baba/ or /baba/ to /bab/ (Lecanuet et al., 1987), and a change in the gender of a speaker, male to female or female to male (Lecanuet, Manera, and Jacquet, 2002) found that the mother speaking elicited a fetal heart rate (FHR) acceleration in both the 2F state (i.e., frequent and periodic body movements, REM, heart rate accelerations during movements, irregular respirations) and the 1F state (i.e., quiescent except for occasional startle, REM absent, heart rate stable with little variability, regular respirations). The tape-recorded voice of the mother or a stranger played via a loudspeaker elicited decelerations in both states. Again, although the response differed depending on whether the mother was speaking or a tape recording was played, no differential responding to mother’s and stranger’s voice was reported. In one other study, DeCasper, Lecanuet, Busnel, Granier-Defere, and Maugerais (1994) found that fetuses could discriminate a tape recording of a familiar versus a novel rhyme read by a female stranger when the familiar rhyme had been repeated by their mother every day from 33 to 37 weeks GA. However, the fetuses failed to respond to an unfamiliar rhyme, which would seem to contradict the findings, noted earlier, that fetuses reliably respond to speech. The findings from these studies leave unanswered the question of whether fetuses are capable of remembering and recognizing their mother’s voice and differentiating between familiar and unfamiliar voices.

There is some evidence from newborn studies to indicate that fetuses may be able to recognize and remember human voices. Using a nonnutritive-sucking paradigm, researchers have demonstrated newborn preference for the mother’s voice over that of a stranger (DeCasper & Fifer, 1980), a story read by the mother during the last 6 weeks of pregnancy (DeCasper & Spence, 1986), and the native language over a foreign language (Moon, Cooper, & Fifer, 1993). The timing of these preferences, occurring shortly after birth, leads to the hypothesis that fetuses are indeed able to remember and recognize human voices that they are exposed to in utero (the in-utero-speech-experience hypothesis). Support for this hypothesis would not only provide evidence of the early onset of speech perception and memory, but would also have implications for theories of the early development of social and cognitive abilities (e.g., language acquisition and attachment).

The present study addressed this hypothesis. Specifically, we used both FHR and body movement measures to index fetal ability to discriminate between maternal and female strangers’ voices. Given that it takes time to perceive the distinct qualities of a voice that distinguish the basis of the acoustic properties of the signals used in the studies. They leave unanswered the question of whether fetal speech discrimination is based on genetic predisposition or influenced by in utero experience with voices.

Evidence of the influence of in utero experience on fetal ability to recognize familiar voices and differentiate between familiar and unfamiliar ones is unclear. To date, few studies have examined whether fetuses are capable of recognizing familiar voices or speech patterns. Hepper, Scott, and Shahidullah (1993) failed to show discrimination of the mother’s and a novel female’s voice, but did show discrimination between the mother speaking and a tape recording of her voice. Similarly, Lecanuet, Manera, and Jacquet (2002) found that the mother speaking elicited a fetal heart rate (FHR) acceleration in both the 2F state (i.e., frequent and periodic body movements, REM, heart rate accelerations during movements, irregular respirations) and the 1F state (i.e., quiescent except for occasional startle, REM absent, heart rate stable with little variability, regular respirations). The tape-recorded voice of the mother or a stranger played via a loudspeaker elicited decelerations in both states. Again, although the response differed depending on whether the mother was speaking or a tape recording was played, no differential responding to mother’s and stranger’s voice was reported. In one other study, DeCasper, Lecanuet, Busnel, Granier-Defere, and Maugerais (1994) found that fetuses could discriminate a tape recording of a familiar versus a novel rhyme read by a female stranger when the familiar rhyme had been repeated by their mother every day from 33 to 37 weeks GA. However, the fetuses failed to respond to an unfamiliar rhyme, which would seem to contradict the findings, noted earlier, that fetuses reliably respond to speech. The findings from these studies leave unanswered the question of whether fetuses are capable of remembering and recognizing their mother’s voice and differentiating between familiar and unfamiliar voices.
it from others, it was necessary to examine fetal behavior over an extended period of time. We made observations of fetal body movements and heart rates during a 2-min period, using a novel passage read by either a familiar or an unfamiliar female adult. The stimuli were delivered at a lower intensity (95 dB SPL) than that used in studies using brief noise presentations to capture fetal attention without startling them. Each tape-recorded mother’s voice served as the familiar voice for her own fetus and the unfamiliar voice for another fetus. This procedure ensured that the acoustic properties of a particular mother’s voice did not account for any differential responses to familiar and unfamiliar voices.

METHOD

Participants

Sixty term fetuses (M = 38.4 weeks GA, SD = 1.1) of Chinese women receiving antenatal care at a hospital in southeast China were tested after the mothers-to-be provided informed verbal consent. All pregnancies were singleton and considered low risk. Testing was conducted on one occasion in a laboratory near the outpatient fetal assessment unit. Gender was not determined at a time of testing.

Equipment

The 2-min speech stimuli were generated by tape-recording mothers reading an adult poem. Maternal speech was recorded and delivered using a Sharp VC-visualized using a Toshiba CAPASEE (Model SSA-220A) real-time

tails, see Coleman, Kisilevsky, & Muir, 1993). Body movements were recording using a digitizer connected to a Macintosh computer (for de-
moments). An FHR for each second was obtained by tracing over the

per strip using a Sonicaid RS232 Cardiotocograph (Oxford Instru-

ments). ND2 SPL meter was used to measure sound intensity.

Continuous FHR was recorded in beats per minute (bpm) on a pa-

er strip using a Sonicaid RS232 Cardiotocograph (Oxford Instru-

ments). An FHR for each second was obtained by tracing over the

recording using a digitizer connected to a Macintosh computer (for de-
tails, see Coleman, Kisilevsky, & Muir, 1993). Body movements were visualized using a Toshiba CAPASEE (Model SSA-220A) real-time ultrasound scanner and video-recorded on-line using a Sharp VC-

RAS Multi Lingual OSD VHS video recorder.

Procedure

During the 6-min procedure, mothers lay on a bed in a semirecum-
bent position. Fetuses were assigned to one of two voice conditions, each of which consisted of three 2-min periods: no stimulus, voice (mother or female stranger), and no stimulus. The same poem was played to the two groups of fetuses; there were 30 fetuses in each group. For the mother’s-voice group, the voice was a tape recording of their own mother’s voice. The voice of a female research team member was used for the first participant in the stranger’s-voice condition, but for every other participant in this group, we played the tape recording that had been played to the previous fetus in the maternal group. The voice stimuli were delivered at an average of 95 dB SPL through a loudspeaker held approximately 10 cm above the maternal abdomen. FHR was recorded continuously, and body movements were video-recorded.

RESULTS AND DISCUSSION

Body Movement

To replicate analyses of body movements in previous studies using short-duration stimulation, we compared the presence of a body move-ment in the first 5 s following voice onset with the presence of a body movement in the 5 s prior to voice onset. This short-term measure did not differ between these two intervals.

Because the stimulus was prolonged in this study, there were a number of other possible longer-term measures. The means and standard deviations for the body movement scores are shown in Table 1.

The total duration of body movements during each period, the number of movements during each period, and the latency to the first movement in each period were compared over the three periods of the study using repeated measures analysis of variance (ANOVA). No significant effects were found.

FHR

The FHR data for all 6 min of recording are shown in Figure 1. To replicate previous work using short-term stimulation, we analyzed heart rate acceleration within the first 10 s after stimulus onset, but found no significant effects. We analyzed the long-term effects of voice on FHR by focusing on three time windows: (a) the 90 s prior to voice onset, (b) the first 90 s following voice onset, and (c) the 90 s following voice offset. We focused on these particular time windows instead of the entire 6 min because the onset and offset of the stimulus could not be regulated with absolute precision. Thus, the data were arranged in two ways: so that the onset of the voice occurred at the same time for each subject, and then again so that the voice offset occurred at the same time. When we analyzed the FHR data for the 90 s prior to voice onset with a 1 (condition: mother vs. stranger; between subjects) × 1 (time: 1–90 s; within subjects) ANOVA, we found no significant change over time, F(89, 5162) = 0.319, p > .05, or condition, F(89, 5162) = 0.882, p > .05. However, when the data for the 90 s following voice onset were analyzed, a difference between the mother and stranger conditions was shown by a significant Time × Voice interaction, F(89, 5162) = 1.31, p < .05. As can be seen in Figure 1, this difference appeared to continue to the end of the recording period, which was confirmed by the analysis of the data for the poststimulus period, which showed only a main effect of voice, F(1, 58) = 4.635, p < .05. The mean FHR was consistently above baseline for fetuses receiving their mother’s voice, but was consistently below baseline for those in the

Table 1. Mean body movement scores in the two conditions

<table>
<thead>
<tr>
<th>Measure and period</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mother’s voice</td>
</tr>
<tr>
<td>Duration of movements</td>
<td></td>
</tr>
<tr>
<td>Prestimulus</td>
<td>6.93 (10.8)</td>
</tr>
<tr>
<td>Stimulus</td>
<td>11.75 (14.1)</td>
</tr>
<tr>
<td>Poststimulus</td>
<td>11.57 (14.6)</td>
</tr>
<tr>
<td>Number of movements</td>
<td></td>
</tr>
<tr>
<td>Prestimulus</td>
<td>1.18 (1.2)</td>
</tr>
<tr>
<td>Stimulus</td>
<td>1.39 (1.6)</td>
</tr>
<tr>
<td>Poststimulus</td>
<td>1.29 (1.3)</td>
</tr>
<tr>
<td>Latency to first movement</td>
<td></td>
</tr>
<tr>
<td>Prestimulus</td>
<td>27.07 (24.9)</td>
</tr>
<tr>
<td>Poststimulus</td>
<td>26.00 (29.2)</td>
</tr>
</tbody>
</table>

Note. Standard deviations are in parentheses.
stranger condition. Prior to stimulus onset, the difference between the mean FHR for the mother’s-voice group and the stranger’s-voice group ranged from −3.06 bpm to 3.16 bpm, but following stimulus onset until the end of the recording period, the mean difference ranged from 2.25 bpm to 8.02 bpm.

To further check the veracity of the results, we looked at the fetuses’ maximum and minimum heart rates during the stimulus and prestimulus control periods. A total of 21 of 30 fetuses in the mother condition had a greater maximum during the stimulus than the control period. This number was significantly greater than would be expected by chance, \( \chi^2(1, N = 30) = 4.80, p < .05 \). Similarly, a total of 21 of 30 fetuses in the stranger condition had a lower minimum during the stimulus than the control period, \( \chi^2(1, N = 30) = 4.80, p < .05 \).

The present results lend support to the in-utero-speech-experience hypothesis. In general, fetuses exposed to their mother’s voice via a loudspeaker showed a 5-bpm increase in heart rate over the first 20 s following voice onset, and this increase was sustained until the end of the recording period. In contrast, when exposed to a female stranger’s voice presented in the same way, fetuses showed a decrease in heart rate of 4 bpm that was similarly sustained. Post hoc analyses determined that the difference in FHR between the groups became significant at 26 s following stimulus onset. This finding cannot be a generalized effect of hearing a voice because the behavioral effects are different.

Also, it cannot be due to the acoustic properties of the passage because the same passages were played to the two groups of fetuses. Although the speaker’s voice varied from fetus to fetus, each female voice was heard twice; the only difference was that the voice was familiar to the fetuses in one group (mother’s voice) and novel to the fetuses in the other (stranger’s voice). Thus, the fetuses’ differential responses to their mother’s and a stranger’s voice suggests that fetuses indeed are capable of remembering and recognizing characteristics of their mother’s voice.

The observation of differential heart rate changes in response to mother’s versus stranger’s voice contrasts with our finding of no differences in fetal body movements. Hepper et al. (1993), who used only body movement measures of responding, also reported no differences in fetal behavioral response to the familiar voice of the mother and the novel voice of a stranger. These disparate findings attest to the need for multiple response measures in studies examining fetal perceptual abilities.

Our results also contrast with those from studies using short bursts of noise at 105 or 110 dB (e.g., Kisilevsky et al., 1989, 2000) or vibroacoustic stimulation (e.g., Kisilevsky, Fearon, & Muir, 1998), in which fetuses showed an immediate, brief body movement coupled with an increase in heart rate within 20 s of stimulus onset. In reviewing the literature on human fetal and newborn behavior, Joseph (2000) related these and other similar behaviors, such as heart rate changes to

---

**Fig. 1.** Average fetal heart rate for the 2 min prior to voice onset, 2 min of mother’s or stranger’s voice, and 2 min following voice offset.
stimuli present throughout gestation. Auditory functioning begins at about the external environment, and also the maternal voice; there are external and includes mother’s heartbeat, intestinal sounds, music and speech from the external world are necessary for the adjustment of the cortical circuits. Excess neurons are pruned, and through dendritic attrition specific neural networks are sculpted and formed (Joseph, 1999), resulting in changes that are relatively permanent and can serve as the basis for subsequent building. In contrast, “experience-dependent” processes refer to the storage of environmental information that is idiosyncratic to the individual; new synaptic connections are formed in response to the events providing the information to be stored. Such learning is open to subsequent change. It seems that experience had to play a role in the differential behavior observed in this study because the same voices were presented to both groups of fetuses, the only difference between groups being the familiarity or novelty of the stimulus. Most likely, experience-expectant and experience-dependent processes do not occur independently, and thus, we cannot say for certain whether the experiential learning in this instance is experience-expectant or experience-dependent. However, we postulate that the repeated exposure to the maternal voice sets up some mechanism that ultimately leads to lasting memories of specific voices—memories that emerge through experience-expectant processes prenatally. Although Joseph (2000) viewed human fetal behaviors as mediated and governed by reflexive brain stem activities, the theory of Greenough et al. (1987) relates perceptual abilities to changes in cortical activity. Much of the support for the experience-expectant phase of development is derived from studies of the visual system (e.g., visual deprivation, ocular dominance; see discussion by Black & Greenough, 1998; Bruer & Greenough, 2001). However, the experience-expectant stage of development cannot occur in the visual system until after birth because of lack of visual stimulation in utero. In contrast, the in utero auditory environment is quite rich and includes mother’s heartbeat, intestinal sounds, music and speech from the external environment, and also the maternal voice; there are external stimuli present throughout gestation. Auditory functioning begins at about 30 weeks GA, so that by term specific neural networks related to the various in utero stimuli may be formed. At birth, infants show a preference for the maternal voice and have the ability to discriminate a wide range of speech stimuli. Their abilities include distinguishing phonemes in both their native language and foreign languages. However, over the course of the 1st year after birth, infants lose this ability for foreign languages (e.g., Werker & Tees, 1984) as a result of lack of exposure. Considering preterm fetuses’ ability to discriminate consonant-vowel sounds (Zimmer et al., 1993), term fetuses’ ability to respond differentially to the mother’s versus a stranger’s voice (reported here), and infants’ change in ability to distinguish phonemes in a foreign language over the 1st year of life, it seems likely that neurological maturation in auditory processing progresses from experience-independent to experience-expectant over the last trimester of pregnancy following the onset of fetal hearing (Kisilevsky et al., 2000), and from experience-expectant to experience-dependent over the 1st year of life. The term “familiar” used throughout this manuscript implies that a sensory stimulus has been experienced frequently over a period of time so that it is recognized (i.e., no longer novel). To recognize something implies learning and memory. If infant recognition of the maternal voice is based on prosodic cues (e.g., Flocchia, Nazzi, & Bertoncini, 2000), then fetuses appear to learn the prosodic information specific to their mother’s voice during repeated exposure, suggesting that some neurological modification based on the maternal voice takes place during fetal maturation. The differences in heart rate exhibited in the current study in response to the familiar maternal voice and the novel female voice suggest memory and learning. In addition, the sustained heart rate changes in response to the voice stimuli indicate that the fetus may have the ability to maintain attention to external stimuli for some time. Such memory and attentional abilities were not reported in any of the literature reviewed by Joseph (2000), and, at this time, we simply do not know where in the brain this memory and learning is occurring. We speculate that such abilities may not be accomplished by the brain stem alone, but may indicate some rudimentary activity at a higher level. There is some support for this notion from preliminary brain-imaging studies. For example, studying cortical activation of the brain in the term fetus using magnetoencephalography, Hynkin et al. (1999) found temporal-lobe activation in response to a tape recording of the mother reciting a nursery rhyme. Moore et al. (2001) replicated and extended this work, noting activation in the frontal lobe as well as the temporal lobes. Thus, our findings suggest that speech processing in the near-term fetus is influenced by experience and may indicate the involvement of rudimentary higher brain functioning.

Our results have several important implications for future research and theorizing about early speech perception and language acquisition. First, the foundation for speech perception and language acquisition may indeed be laid before birth, as suggested by theorists such as Kuhl (1988) and Lindblom (1992). Second, the precocious language-processing abilities observed in newborns and young infants may not be due to a hard-wired speech-processing module in the brain, as assumed by Chomsky (1980) or Liberman and Mattingly (1985); rather, in utero experience may play a critical role. Finally, at least in the domain of speech processing, there may be signs of perception, memory, and attention suggesting some involvement of higher brain structures in the prenatal period.

Acknowledgments—This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada to B.S. Kisilevsky and K. Lee, as well as a summer studentship from the Canadian Institutes for Health Research to S. Schmidt in Kisilevsky’s laboratory. Portions of this article were presented at the International Conference on Infant Studies, Toronto, Ontario, Canada, in April 2002.
REFERENCES


