

## The Tuning of Human Neonates' Preference for Speech

Athena Vouloumanos  
*New York University*

Marc D. Hauser  
*Harvard University*

Janet F. Werker  
*University of British Columbia*

Alia Martin  
*New York University*

Human neonates prefer listening to speech compared to many nonspeech sounds, suggesting that humans are born with a bias for speech. However, neonates' preference may derive from properties of speech that are not unique but instead are shared with the vocalizations of other species. To test this, thirty neonates and sixteen 3-month-olds were presented with nonsense speech and rhesus monkey vocalizations. Neonates showed no preference for speech over rhesus vocalizations but showed a preference for both these sounds over synthetic sounds. In contrast, 3-month-olds preferred speech to rhesus vocalizations. Neonates' initial biases minimally include speech and monkey vocalizations. These listening preferences are sharpened over 3 months, yielding a species-specific preference for speech, paralleling findings on infant face perception.

How do initial perceptual biases and learning from specific experience interact to prepare infants to adapt optimally to their world? Across perceptual domains, infants use early biases to organize the world. For example, young infants are able to discriminate non-native speech sounds that are unfamiliar to them (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984) and identify individual faces of species they have never encountered before (Pascalis, de Haan, & Nelson, 2002). By the end of the 1st year of life, these biases become attuned to sounds and visual information in the infant's environment, such that infants no longer readily discriminate non-native speech sounds (Kuhl et al., 1992; Werker & Tees, 1984) nor do they identify unfamiliar species' faces (Pascalis et al., 2002). Initial biases, and rapid

attunement, may constitute conserved and fundamental principles that underwrite the development of perceptual systems across the animal kingdom. In this article, we explore the ontogeny of the specificity of the neonatal preference for human speech.

Many species are endowed with an innate preference to attend to the visual and auditory cues of their conspecifics (Gottlieb, 1997; Johnson, Bolhuis, & Horn, 1985; Marler & Peters, 1977). For example, precocial birds preferentially approach the calls of conspecifics, whereas the young fledglings of song learning birds will selectively acquire their species-specific song if they are exposed to it during a sensitive period (Gottlieb, 1997; Marler & Peters, 1977). A common perspective in studies of language acquisition is that humans are similarly endowed with a preference for conspecific vocalizations and, more specifically, that speech is special for humans (Liberman & Whalen, 2000; Pinker & Jackendoff, 2005), perceived in unique ways from the earliest stages of development (Vouloumanos & Werker, 2007). Support for this hypothesis comes from studies showing that neonates exhibit a preference for speech over filtered (Spence & DeCasper, 1987) and reversed speech (Peña et al., 2003), as well as synthesized signals such as complex synthetic sounds (Vouloumanos & Werker, 2007) and white noise

---

Funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Centre for Research in Language, Mind and Brain, and the Fonds québécois de recherche sur la société et la culture to A. V.; by the James S. McDonnell Foundation and 21st Century grants to M. D. H.; by NSERC, the Canada Research Chair Program, and the James S. McDonnell Foundation to J. F. W., and by an NSERC Undergraduate Summer Research Award to A. M. M. D. H.'s research was also funded by gifts from J. Epstein and S. Shuman. We thank Erin Moon, Marisa Cruickshank, and Vivian Pan for help with data collection. For comments on earlier versions of the manuscript, we thank Steven Pinker and especially Gary Marcus. We are especially grateful to participating parents and infants.

Correspondence concerning this article should be addressed to Athena Vouloumanos, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. Electronic mail may be sent to athena.vouloumanos@nyu.edu.

(Butterfield & Siperstein, 1970). Such a preference for speech suggests that infants are prepared to attend to auditory stimuli in their environment that are relevant for language, thus facilitating language acquisition (Kuhl, Coffey-Corina, Padden, & Dawson, 2005).

Although previous studies support the hypothesis that speech is a privileged sound for infants, they leave open the possibility that neonates' listening preference is based on properties of speech that are not unique to speech but instead are shared with the vocalizations of related species. Speech and nonhuman primate vocalizations share some acoustic properties and are coded similarly in the human auditory system (Altmann, Doehrmann, & Kaiser, 2007; Smith & Lewicki, 2006). We selected rhesus monkey calls for comparison because, although acoustically distinct from human speech, they are similarly spectrally rich. In particular, rhesus have a broad vocal repertoire spanning a wide range of activities, including calls produced during the discovery of food sources (e.g., coos, warbles) and affiliative social behavior (e.g., girneys; Hauser, 2000). Like human speech, some rhesus monkey calls are produced by oscillations of the larynx (phonation) and filtered by the supralaryngeal vocal tract, which dampens or amplifies different frequencies (Fitch & Hauser, 1995; Hauser, Evans, & Marler, 1993; Rendall, Owren, & Rodman, 1998; see Figure 1). This filtering of frequencies generated by the laryngeal oscillations produces discrete formants, or concentrations of energy, whose distribution and fluctuations across time generate a wide range of speech sounds. As a result of constraints on the movements of rhesus articulators, however, the "phonetic" repertoire of rhesus monkeys is limited by comparison with that of humans (Belin, 2006; Fitch, 2000) and the frequency spectrum of calls is broader and higher so that, perceptually, rhesus calls present a broader distribution of spectral

energy and a higher pitch than human speech (Fitch, 1997; Owren & Rendall, 2003). Although some rhesus calls have an acoustic resemblance to human affective vocalizations, adult human listeners do not perceive rhesus calls as being generated by humans (see ratings in the Method section that follows), nor do they explicitly distinguish these calls based on their affective valence (Belin et al., 2008). Testing a preference for speech against the vocalizations of a closely related primate provides a strong test for an early specialization for speech in humans.

Using a contingency procedure in which infant participants could control the presentation of different sounds, we tested neonates on their preference for speech compared to the vocalizations of another primate—the rhesus macaque (*Macaca mulatta*). There are three possible outcomes. First, if neonates prefer speech to rhesus calls, then this supports earlier reports that speech is privileged for human neonates, distinct from the vocalizations of another related species. The neonatal preference for speech would then be based on properties that are not shared with rhesus calls. Second, if neonates prefer rhesus calls to speech, then the task is most likely picking up on a novelty preference. Although this would be a surprising outcome as previous studies showed a neonatal preference for speech over many other novel sounds, previous studies never evaluated the preference for human speech against the calls of a related species. Finally, if neonates show equal interest in speech and rhesus calls, with no preference for one sound over the other, then this suggests that the initial bias for speech may derive from properties shared by speech and nonhuman primate vocalizations. In this case, to ensure that the lack of preference was not simply the result of a methodological failure, it would be important to show—as in earlier work on the neonatal speech preference—that neonates also prefer rhesus vocalizations to other acoustic signals. Finally, we

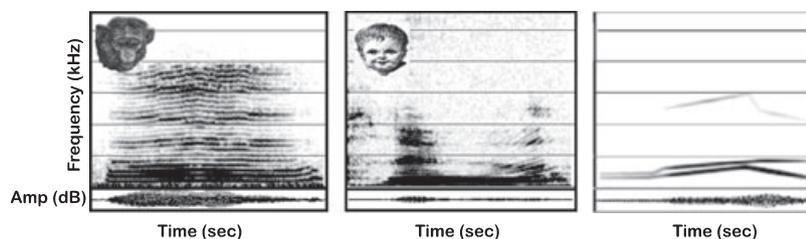


Figure 1. Spectrogram (top) and waveform (bottom) of a rhesus monkey vocalization (coo; on the left), a human speech token ("plooo"; in the middle), and a synthetic nonspeech sound (on the right).  
Note. Horizontal lines denote frequency increments of 2000 Hz.

examine the listening preferences of 3-month-old infants.

### Experiment 1

We tested human neonates on their preference for speech compared to rhesus calls using an infant-controlled high-amplitude (HA) sucking method (e.g., Butterfield & Siperstein, 1970; DeCasper & Fifer, 1980; Moon, Cooper, & Fifer, 1993; Vouloumanos & Werker, 2007). In this procedure, neonates sucked on a sterilized non-nutritive pacifier to control the presentation of sounds. We presented nonsense speech using English-like speech syllables and rhesus monkey calls in alternating trials, and measured the number of neonates' HA sucks for each of the two types of vocalizations.

#### Method

##### Participants

Thirty full-term (minimum of 37.5 weeks gestational age) human neonates ( $M = 47$  hr, range = 18–120 hr; 12 females and 18 males) completed the study. An additional 44 neonates were excluded for noncompliant behavior, such as fussing or falling asleep (24), parental interference (1), or failing to meet established experimental criteria (Floccia, Nazzi, & Bertoni, 2000; Vouloumanos & Werker, 2007) of a minimum of 10 HA sucks in the first 4 min and at least 1 suck in every subsequent experimental minute (19). Neonates were tested at a birthing Hospital in Vancouver, British Columbia. Reflecting the diversity of the local community, the languages spoken in neonates' home environment varied, with the percentage of English spoken between 0% and 100%. We did not collect additional demographic information about family ethnicity. Parents gave informed consent on behalf of their infants. All procedures were approved by the university's Research Ethics Board.

##### Stimuli

In selecting specific monkey and human stimuli, it was important to control for factors that in and of themselves might bias infants toward one or the other group of sounds. As neonates listen less to angry and sad voices than they do to happy or neutral voices (Aldridge, 1994, as cited in Walker-Andrews, 1997), we excluded rhesus calls produced during aggressive behaviors (e.g., screams, barks,

and copulatory screams). Instead, we selected calls produced during positive social contexts such as affiliative interactions (girneys, coos) and the discovery of food (coos, warbles; Hauser, 2000) considered to be of positive affect in previous studies (Belin et al., 2008; Nicastro & Owren, 2003).

*Speech.* Ten nonsense speech sounds were produced in infant-directed speech by three native English female speakers (Speaker 1: "plooo," "keev," "yut," "boola," "nahod," "kraw"; Speaker 2: "trom," "fi"; Speaker 3: "dup," "makalak") with average pitch of 283 Hz, and average duration of 468 ms.

*Monkey calls.* Ten rhesus monkey calls were recorded on Cayo Santiago, Puerto Rico, from three female rhesus monkeys (Caller 1: 3 coos [social or food related], 3 girneys [affiliative]; Caller 2: 1 coo, 1 warble [food related]; Caller 3: 1 coo, 1 girney) with average pitch of 346 Hz and average duration of 531 ms. Calls were recorded with a highly directional shotgun microphone at a distance of 1–3 m. The calls were selected such that there was a high signal-to-noise ratio, only one concurrent vocalization, and no interference from other biotic sounds. Stimulus duration did not differ,  $t(18) < 1$ ,  $ns$ , but rhesus calls naturally exhibit a slightly higher pitch,  $t(17) = 2.3$ ,  $p < .05$ . (For description of acoustic morphology and context, see Hauser, 1992, 1996; Hauser & Fowler, 1992; Hauser & Marler, 1993.)

*Adult ratings of vocalizations.* Thirty undergraduates rated the 20 primate vocalizations described on a 5-point scale for how likely they were to have been produced by a human. The 10 monkey calls were rated as significantly less human-like than the 10 human nonsense speech ( $M = 2.97$ ,  $SE = 0.924$  for monkey vocalizations;  $M = 4.63$ ,  $SE = 0.122$  for human nonsense speech),  $t(18) = 5.61$ ,  $p < .001$ .

##### Apparatus and Procedure

Neonatal listening preferences were quantified by means of a HA sucking procedure (Eimas, Siqueland, Jusczyk, & Vigorito, 1971) in which the presentation of different sounds was contingent on the newborn delivering an HA suck (Vouloumanos & Werker, 2007). Approximately 2 hr after feeding, neonates were brought to a sound-attenuated room and placed supine in a bassinette. During the experiment, a parent was seated behind and out of reach of the bassinette. An experimenter wearing headphones playing music presented neonates with a sterilized pacifier coupled to a Gould P23

pressure transducer, which was connected to an IBM-PC-compatible computer. After presenting the pacifier to the neonate, the experimenter only interacted with the newborn to reinsert the pacifier if it had been released. If the pacifier could not be reinserted immediately, the neonate's data were discarded from the sample. The computer registered air pressure changes caused by infant sucks.

Each individual infant's sucking amplitude range was measured during a silent baseline minute. The mean baseline HA sucking rate was 33.7 HA sucks/min ( $SE = 2.93$ ). Following the calculation of their baseline sucking amplitude, neonates were presented with a sound stimulus whenever they applied suction in the upper 80% of their sucking amplitude range. The stimulus onset asynchrony (from the start of one stimulus to that of the next) was set at 1500 ms to accommodate the longest sound duration. The presentation rate was thus one stimulus per 1.5 s. All sounds were presented at 70 dB ( $\pm 4$  dB). The presentation of speech and rhesus calls was alternated every minute, with order counterbalanced across participants, and the 8 min postbaseline submitted to analysis. The HA sucking software implements the minute-by-minute change independently of the baby's sucking behavior as in many HA sucking experiments (Bertoncini, Bijeljac-Babic, Blumstein, & Mehler, 1987; Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Floccia et al., 2000; Mehler et al., 1988; Nazzi, Floccia, & Bertoncini, 1998). To ensure that newborns heard sounds during every minute, we excluded newborns that produced fewer than 10 sucks in the first four experimental minutes or 0 sucks during any of the 8 min. Within each minute, tokens were presented randomly with replacement allowing repetitions of a given token. Earlier applications of this procedure have been successful in showing that newborns prefer speech over complex synthetic sounds (Vouloumanos & Werker, 2007).

### Results and Discussion

As there were no significant effects of infant sex or sound order on listening preferences ( $F < 1$ , *ns*), we collapsed across these factors for all analyses. As sampling over blocks can yield more stable measurements and has successfully revealed neonatal preferences in previous studies (Vouloumanos & Werker, 2007), we analyzed HA sucks in two observation blocks (first 4-min block vs. second 4-min block) in a repeated measures analysis of variance (ANOVA). Results indicated no main effects and no interactions (all *ns*). Neonates sucked equiva-

lently for speech ( $M = 33.3$  HA sucks,  $SE = 2.4$ ) and for rhesus calls ( $M = 33.8$  HA sucks,  $SE = 2.7$ ),  $F(1, 29) < 1$ , *ns*. An additional ANOVA analyzing preference for speech and monkey calls for each experimental minute also revealed no main effects and no interactions (all *ns*). Next, we examined whether neonates' HA sucking changed over the two observation blocks for speech or for rhesus calls and found no statistically significant effect (using Bonferroni corrections for multiple comparisons) for either speech,  $t(29) = 2.12$ , *ns*, or rhesus calls  $t(29) = 1.02$ , *ns* (see Table 1). Thirteen of 30 neonates exhibited a preference for speech (see Figure 2). To examine whether the amount of English exposure in the neonates' home environment had an effect on neonates' listening preferences, we ran an ANOVA with sound type (speech, rhesus) and percentage of English exposure (as estimated by parental report at the time of consent) as a covariate and found no main effect of English exposure,  $F(1, 28) < 1$ , *ns*, and no modulatory effect of English exposure on neonates' listening preference for speech and rhesus calls,  $F(1, 28) = 1.19$ , *ns*.

Table 1

*High-Amplitude (HA) Sucking Across the Three Studies: (a) Speech Versus Rhesus Monkey Calls, (b) Rhesus Monkey Calls Versus Synthetic Sounds, and (c) Speech Versus Synthetic Sounds (Reported in Vouloumanos & Werker, 2007)*

| Sound comparison                            | <i>n</i> | Sound            | HA Sucking Block 1 | HA Sucking Block 2 | Relative change in HA sucking |
|---|----------|------------------|--------------------|--------------------|-------------------------------|
| Speech versus monkey calls                  | 30       | Speech           | 71.20              | 61.73              | -4.4                          |
|   |          | Monkey calls     | 69.90              | 64.83              |                               |
| Monkey calls versus synthetic sounds        | 16       | Monkey calls     | 68.75              | 60.50              | 12.3*                         |
|   |          | Synthetic sounds | 74.50              | 54.00*             |                               |
| Speech versus synthetic sounds <sup>a</sup> | 22       | Speech           | 71.91              | 73.00              | 14.5*                         |
|   |          | Synthetic sounds | 76.91              | 63.45*             |                               |

<sup>a</sup>Previously reported in Vouloumanos and Werker (2007).

*Note.* Neonates' HA sucking decreases significantly across blocks for the synthetic sounds; no significant decrease across blocks is observed for either speech or rhesus monkey calls. The relative change in HA sucking in the speech versus rhesus call comparison is significantly different both from the monkey call versus synthetic sound comparison, and from the speech versus synthetic sound comparison. The relative change in HA sucking for monkey over synthetic sounds is equal to that for speech over synthetic sounds. See the Results section for more details.

\* $p < .05$ .

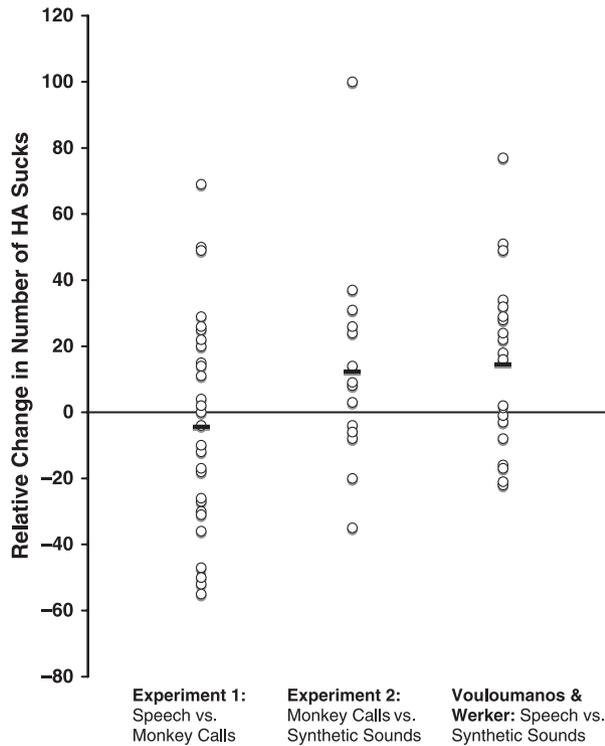


Figure 2. Scatterplot of individual neonates' relative change in the number of high-amplitude sucks for sound types across the three experiments.

Note. For Experiment 1, we plot as positive the relative change for speech over monkey calls; for Experiment 2, we plot as positive the relative change for monkey calls over synthetic sounds; and for Vouloumanos and Werker (2007), we plot as positive the relative change for speech over synthetic sounds.

Although earlier work showed that neonates preferred speech to artificial sounds and manipulated speech sounds (Butterfield & Siperstein, 1970; Peña et al., 2003; Spence & DeCasper, 1987; Vouloumanos & Werker, 2007), this study demonstrates that neonates have no preference for speech over rhesus calls, responding equivalently to both.

## Experiment 2

To ensure that the equivalent interest in speech and rhesus calls reflects a legitimate preference for listening to these two types of vocalizations, as opposed to a null result, we ran an additional condition. In a previous study, neonates preferred speech to complex synthetic sounds (Vouloumanos & Werker, 2007). These complex synthetic sounds were composed of four sinusoidal waves that emulated the frequency changes of speech and thus contained a similar pitch contour and some of its

spectral features (Vouloumanos & Werker, 2004, 2007); in some contexts, adult listeners can even perceive these types of complex synthetic sounds as speech-like (Remez, Rubin, Pisoni, & Carrell, 1981). In this previous study, neonates manifested a preference for speech by adjusting their sucking rate over the course of the experiment so as to listen to speech more than to the complex synthetic sounds. Over the two observation blocks, neonates in this previous study maintained their HA sucking rate for speech,  $t(21) = 0.315$ , *ns*, but significantly decreased their sucking for the complex synthetic sounds,  $t(21) = 2.92$ ,  $p < .02$  (Vouloumanos & Werker, 2007). To test whether human neonates show a similar sustained interest in listening to rhesus monkey calls, we presented rhesus monkey calls and complex synthetic sounds to 16 neonates. If both speech and rhesus calls are preferred vocalizations for human neonates, then neonates' HA sucking in this experiment should show differential arousal to rhesus monkey calls over complex synthetic sounds.

## Method

### Participants

Sixteen full-term (minimum of 37.5 weeks gestational age) human neonates ( $M = 46$  h, range = 20–97 hr; 7 females and 9 males) completed the study. An additional 21 neonates were excluded for fussing or falling asleep (10), parental interference (1), or not meeting the established experimental criteria (10). Neonates were tested at a birthing hospital in Vancouver, British Columbia. As in Experiment 1, the percentage of English spoken in neonates' home environment varied between 0% and 100%, and additional ethnographic information was not collected. Parents gave informed consent on behalf of their infants. All procedures were approved by the university's Research Ethics Board.

### Stimuli

*Monkey calls.* These were identical to those used in Experiment 1. As there were four different complex synthetic tokens in total, we randomly selected a different subset of 4 monkey tokens from the full set of 10 tokens for presentation for each neonate.

*Synthetic sounds.* Synthetic sounds consisted of time-varying sinusoidal waves tracking the main regions of significant energy, namely, the fundamental frequency and the first three formants of a nonsense speech token. By tracking changes across

time for the peak frequencies of their speech counterparts, these synthetic sounds follow very closely the spectral and timing changes of speech. The average pitch of the synthetic sounds was 227 Hz, and the average duration was 589 ms. Stimulus pitch did not differ between sound types,  $t(8) = 2.2$ , *ns*, but synthetic sounds were significantly longer than rhesus monkey calls,  $t(8) = 4.1$ ,  $p < .01$  (for a complete description of the stimuli, please see Vouloumanos & Werker, 2004). Consistent with previous studies with similar synthetic sounds (Remez & Rubin, 1993), adults in pilot studies did not report perceiving these synthetic sounds as speech.

#### *Apparatus and Procedure*

These were identical to Experiment 1. The presentation of rhesus calls and synthetic tokens was alternated every minute, with order counterbalanced across participants, and the 8 min postbaseline submitted to analysis. Within each minute, tokens were presented randomly with replacement.

#### *Results and Discussion*

Neonates preferentially listened to rhesus monkey calls, maintaining their HA sucking for rhesus calls,  $t(15) = 2.01$ , *ns*, while decreasing their HA sucking for complex synthetic sounds,  $t(15) = 2.64$ ,  $p = .038$ , over the two observation blocks. Eleven of the 16 neonates exhibited a preference for rhesus monkey calls (see Figure 2). As a final direct test of listening preferences, we compared the relative change in HA sucking across the three studies: Speech versus rhesus calls (Experiment 1), speech versus complex synthetic sounds (Vouloumanos & Werker, 2007), and rhesus calls versus complex synthetic sounds (Experiment 2). In planned comparisons, we found that the HA sucking pattern of neonates' privileging rhesus calls over complex synthetic calls was similar to that of neonates privileging speech over complex synthetic calls (one-tailed, *ns*), and different from neonates that listened to speech and monkey calls equivalently (one-tailed,  $p < .05$ ; see Table 1).

Neonates responded preferentially to speech and to rhesus calls over complex synthetic sounds but showed no preference for speech over rhesus calls. Together, these studies suggest that neonates privilege both speech and calls of rhesus monkeys. The auditory biases of human neonates appear more broadly tuned to properties that are shared by the vocalizations of nonhuman primates.

### **Experiment 3**

To examine whether infants' listening preferences change during early development, we tested 3-month-olds on their preference for speech over rhesus calls. Three-month-old infants were presented with nonsense speech sounds and rhesus monkey calls in alternating trials using an infant-controlled sequential looking preference procedure (Cooper & Aslin, 1990; Vouloumanos & Werker, 2004) in which infants could use their looking behavior to control the presentation of sounds. We measured infants' average looking times for the two types of vocalizations.

#### *Method*

##### *Participants*

Sixteen 3-month-old infants ( $M = 100$  days, age range = 87–111 days) completed the study. An additional 20 infants were excluded as a result of fussing or crying (6), inattention to the screen (5), difficulty seeing infants' eyes (2), failure to look away from the screen to end the trial before the maximum trial length of 60 s had been reached for three or more of the test trials (2), or experimenter error (5). Parents gave informed consent on behalf of their infants. All procedures had been approved by the university's Research Ethics Board.

##### *Stimuli*

The auditory stimuli were the same 10 human nonsense speech tokens and 10 rhesus monkey calls that were used in Experiment 1.

##### *Materials and Procedure*

Infants were tested using an infant-controlled sequential looking preference procedure (Cooper & Aslin, 1990; Vouloumanos & Werker, 2004) in which infants controlled trial onset and offset by looking at or away from a monitor displaying a fixation checkerboard. Testing was conducted in a sound-attenuated room where infants were seated on a parent's lap in front of a computer monitor. Parents wore headphones playing music to mask the presentation of the experimental sounds. Sounds were played with an average amplitude of 60 dB ( $\pm 5$  dB) from speakers placed below the monitor. Following a musical pretest, infants heard speech and rhesus calls in alternating trials, with order counterbalanced across participants, for a total of eight trials, four trials per sound type.

Infants' attention was drawn to the screen by a red flashing light at the beginning of the experiment and between every trial. Once the infant fixated on the screen, a black and white checkerboard was presented in tandem with one of the sets of sounds. Presentation ceased when the infant looked away from the monitor for a consecutive 2-s period or the maximum trial length of 60 s had been reached. Once the infant looked back at the screen, presentation of the checkerboard recommenced with a new set of sounds. Within each trial, tokens were presented in semirandom order, such that no tokens were ever repeated consecutively. All infants heard a Bach violin concerto in a pretest before the experimental trials began to familiarize them with the procedure. Infants' average looking times for human speech and rhesus monkey calls were calculated based on frame-by-frame offline coding (30 frames/s) of infant looks toward the monitor during each test trial by a coder blind to the experimental condition being tested. Reliability coding performed on 25% of the trials was extremely high (Pearson's  $R = .999$ ,  $p < .01$ ). Three infants were missing values for one trial within a set of four human or four monkey vocalizations; these values were replaced by taking an average of the remaining 15 infants' looking times for that particular sound trial.

#### Results and Discussion

As there were no significant main or interactive effects of infant sex or order on looking time ( $F < 1$ ,  $ns$ ), we collapsed across these between-subject factors and performed a repeated measures ANOVA on infant looking time with sound type (speech, monkey calls) and trials (1, 2, 3, 4) as within-subject factors. Results revealed a significant main effect of sound type, with 3-month-olds looking longer for speech ( $M = 14.6$  s,  $SE = 2.01$ ) compared with rhesus calls ( $M = 11.5$  s,  $SE = 1.67$ ),  $F(1, 15) = 8.49$ ,  $p = .01$ , as well as a main effect of trial,  $F(3, 45) = 5.13$ ,  $p < .01$ , with overall looking times decreasing in later trials. Twelve of the sixteen 3-month-olds looked longer for speech, as compared to 13 of 30 neonates, Pearson's chi-square  $\chi^2(1, N = 46) = 4.22$ ,  $p < .05$  (see Figure 3), confirming a developmental difference in infants' preference for speech relative to rhesus monkey calls.

By at least the age of 3 months, human infants prefer listening to speech compared to rhesus calls. Rapidly in early human development, the relatively broad tuning of neonatal listening preferences is sharpened, becoming selective for speech.

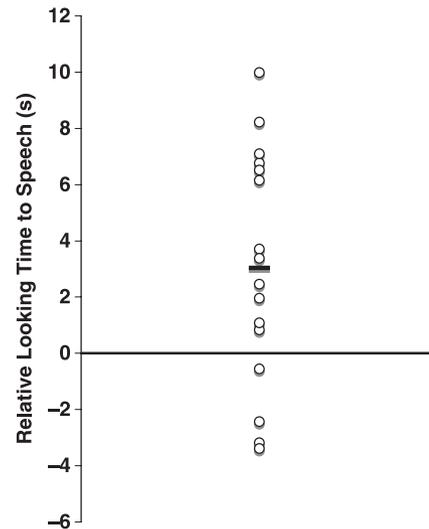


Figure 3. Scatterplot of individual 3-month-olds' relative looking time for speech compared to rhesus monkey vocalizations.

Note. We plot as positive the relative change for speech over monkey calls.

#### General Discussion

In contrast to previous studies suggesting that humans are born with a listening preference for speech (Vouloumanos & Werker, 2007), these results demonstrate that neonates' listening biases privilege human speech and the calls of another primate, the rhesus macaque. Infants' initial, relatively broad, tuning sharpens rapidly over the course of the first 3 months, resulting in a preference for speech over rhesus calls.

The lack of preference for speech over rhesus monkey calls in neonates is striking, especially when considered in light of previous studies on auditory preferences in human neonates. Neonates typically show preferences for specific speech sounds heard prenatally: They prefer listening to their mother's voice compared to the voice of other females (DeCasper & Fifer, 1980), privilege the rhythmic pattern of their native language over that of unfamiliar languages (Mehler et al., 1988; Moon et al., 1993), and show postnatal recognition of specific songs and stories that had been read to them prenatally (DeCasper & Spence, 1986). These prior findings indicate that the human womb is permeable to sound and, crucially, that the auditory system of the fetus can perceive, encode, and remember some of the specific features of the sounds that are available in the prenatal environment. Given infants' prenatal exposure to speech and their capacity to hear in the third trimester, as well as our selection of speech sounds with attractive intonation patterns (Cooper, Abraham, Ber-

man, & Staska, 1997; Fernald, 1985), one might have expected neonates to prefer speech to monkey calls. Instead, the comparison between speech and monkey calls—as well as the experimental condition confirming a preference for rhesus monkey calls over complex synthetic sounds—demonstrates that speech and rhesus monkey calls can be similarly engaging to newborns.

Why do neonates fail to show a preference for relatively more familiar speech sounds when these are compared to rhesus monkey calls given that they recognize and prefer familiar native and maternal speech vocalizations when various *speech* sounds are contrasted? Previous speech preference studies with newborns used full sentences composed of connected speech (DeCasper & Fifer, 1980; Mehler et al., 1988; Moon et al., 1993). In contrast, our preference studies used isolated words (as in Vouloumanos & Werker, 2004, 2007). Compared to single nonsense words, the connected speech of sentences is longer and, partly as a function of greater length, conveys more of the acoustic (e.g., syllable length, formant structure) and linguistic (e.g., rhythmic, intonational) properties of speech. Sentences may thus allow neonates to extract more information about the signal, including acoustic cues to maternal voice and native language. A comparison of connected speech and rhesus monkey vocalizations could thus reveal a neonatal preference for speech, but isolated words do not.

Neonates may have a general interest in the acoustic properties of voices, harmonically rich sounds composed of quasiperiodic waves punctuated by spectral and temporal irregularities (e.g., Belin, 2006). If this were the case, any vocalization produced by a range of species may engage newborns. More narrowly, neonatal listening preferences may extend to acoustic properties that are shared by other nonhuman primates or may be restricted to vocalizations that show more human-like spectral filtering. The vocal repertoire of rhesus monkeys is certainly not unique among nonhuman primates, as many other monkeys and apes demonstrate both spectral and temporal variation, with resulting differences in affective and semantic information (Arnold & Zuberbuhler, 2006; Owren, Seyfarth, & Cheney, 1997; Rendall, Seyfarth, Cheney, & Owren, 1999; Zuberbuhler, Cheney, & Seyfarth, 1999). More importantly for this study, rhesus monkeys and humans show similarities in the production mechanisms of their vocalizations. Like humans, rhesus monkeys may be able to control the laryngeal source of their calls independently of the supralaryngeal filter, suggesting possible

similarities between human and rhesus neuromotor organization (Hauser & Schon-Ybarra, 1994; Hauser et al., 1993; Rendall et al., 1998). Although few studies have directly examined how humans perceive rhesus monkey calls (cf. Belin et al., 2008; Lewkowicz & Ghazanfar, 2006; Owren & Rendall, 2003), the superficial similarities between the two species' vocalizations may have been sufficient to engage neonates.

What accounts for the developmental shift in infant listening preferences between birth and 3 months? One possibility is that newborns simply could not discriminate between speech and rhesus monkey calls, whereas 3-month-olds did discriminate and therefore preferred listening to speech. Although this possibility cannot be ruled out, it seems unlikely because newborns robustly discriminate much more subtle acoustic differences, for example, between [ba] and [da] (Bertoncini et al., 1987) and between [taep] and [paet] (Moon, Bever, & Fifer, 1992). Our results are thus most likely explained by neonates' engagement with the sounds rather than a failure to discriminate.

A second possibility is that 3-month-olds' preference for speech is a product of increased experience with specific properties of English. Although the increased experience is undeniable, other studies have shown that infants do not recognize native language sound sequences or stress until 9 months of age (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Friederici, Wessels, & Svenkerud, 1993) and, crucially, do not become attuned to the specific speech sounds of their native language until 6–10 months (Kuhl et al., 1992; Werker & Tees, 1984). Given this relatively late tuning (6–10 months) to native language sound structure, it seems unlikely that the speech preference observed in these relatively younger 3-month-olds in our studies derived specifically from properties of English. Instead, accrued exposure to speech of any kind, or human vocalizations in general, is more likely to contribute to the sharpening of infant preferences.

Finally, the developmental shift may also reflect a shift in infants' attention to different properties of sounds. By 3 months, infants may start to attend to some characteristic properties of speech, such as the alternations between rapid frequency transitions and steady-state frequencies that characterize consonant–vowel alternations (Ladefoged, 2006). Concurrently or alternatively, infants may attend to the specific frequency range (50–5000 Hz) that is characteristic of the human voice. In this latter case, neutral and positively valenced vocalizations that are recognized by infants as human productions,

including yawns and laughter, would trigger infants' attention.

The observed shift from more general to species-specific processing in the vocal domain parallels recent results on infant face processing and cross-modal matching: Infants begin life without a human-specific face bias but by 9 months show better processing of human faces compared to the faces of other species (Lewkowicz & Ghazanfar, 2006; Pascalis et al., 2002). Such tuning mechanisms have been observed not only for faces but for language as well. Infants' initial language-general discrimination abilities become language specific by the end of their 1st year (Kuhl et al., 1992; Werker & Tees, 1984). It seems likely that tuning mechanisms are at work in early infancy, sharpening broad initial biases into more specific ones in many domains, and that this may represent a general principle of early perceptual development (Scott, Pascalis, & Nelson, 2007).

Auditory tuning to conspecifics, observed here by 3 months, appears earlier than visual tuning to conspecifics, apparent only at 9 months (Lewkowicz & Ghazanfar, 2006; Pascalis et al., 2002). This is likely because of a combination of maturational and experiential differences in the two systems. Specifically, whereas the auditory system appears functional and responsive by the 6th month of gestation (Gerhardt & Abrams, 2000; Kisilevsky et al., 2003), the visual system, although functional before birth, can only respond to external visual stimuli such as faces after birth. Humans become tuned to conspecific auditory stimuli, and speech in particular, earlier and on a faster time scale (6 vs. 9 months of relevant experience) than they become tuned to conspecific faces. This may very well be because speech, the communicative signal of human conspecifics, is special, and rapid tuning reflects this special status.

## References

- Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex*, *17*, 2601–2608.
- Arnold, K., & Zuberbuhler, K. (2006). Language evolution: Semantic combinations in primate calls. *Nature*, *441*(7091), 303.
- Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical transactions of the Royal Society of London Series B: Biological Sciences*, *361*(1476), 2091–2107.
- Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proceedings of the Royal Society of London Series B, Biological Sciences*, *275*(1634), 473–481.
- Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. E., & Mehler, J. (1987). Discrimination in neonates of very short CVs. *Journal of the Acoustical Society of America*, *82*, 31–37.
- Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, *117*, 21–33.
- Butterfield, E. C., & Siperstein, G. N. (1970). Influence of contingent auditory stimulation upon non-nutritional suckle. In J. F. Bosma (Ed.), *Third symposium on oral sensation and perception: The mouth of the infant* (pp. 313–334). Springfield, IL: Charles C. Thomas.
- Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior & Development*, *20*, 477–488.
- Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.
- DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, *9*, 133–150.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(968), 303–306.
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development*, *8*, 181–195.
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America*, *102*, 1213–1222.
- Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*, 258–267.
- Fitch, W., & Hauser, M. D. (1995). Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on "honest" advertisement. *American Journal of Primatology*, *37*, 191–219.
- Floccia, C., Nazzi, T., & Bertoncini, J. (2000). Unfamiliar voice discrimination for short stimuli in newborns. *Developmental Science*, *3*, 333–343.
- Gerhardt, K. J., & Abrams, R. M. (2000). Fetal exposures to sound and vibroacoustic stimulation. *Journal of Perinatology*, *20*(8, Pt. 2), S21–S30.
- Gottlieb, G. (1997). *Synthesizing nature-nurture: Prenatal roots of instinctive behavior*. Mahwah, NJ: Erlbaum.
- Hauser, M. D. (1992). Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: A learned mode of production? *Journal of the Acoustical Society of America*, *91*, 2175–2179.
- Hauser, M. D. (1996). Vocal communication in *Macaques*. In J. E. Fa & D. G. Lindburg (Eds.), *Evolution and ecology of macaque societies* (pp. xvii). Cambridge, UK: Cambridge University Press.

- Hauser, M. D. (2000). A primate dictionary? Decoding the function and meaning of another species' vocalizations. *Cognitive Science*, *24*, 445–475.
- Hauser, M. D., Evans, C. S., & Marler, P. (1993). The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. *Animal Behaviour*, *45*, 423–433.
- Hauser, M. D., & Fowler, C. A. (1992). Fundamental frequency declination is not unique to human speech: Evidence from nonhuman primates. *Journal of the Acoustical Society of America*, *91*, 363–369.
- Hauser, M. D., & Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behavioral Ecology*, *4*, 194–205.
- Hauser, M. D., & Schon-Ybarra, M. (1994). The role of lip configuration in monkey vocalizations: Experiments using xylocaine as a nerve block. *Brain and Language*, *46*, 232–244.
- Johnson, M. H., Bolhuis, J. J., & Horn, G. (1985). Interaction between acquired preferences and developing predispositions during imprinting. *Animal Behaviour*, *33*, 1000–1006.
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*, 675–687.
- Jusczyk, P. W., Friederici, A. D., Wessels, J. M., & Svenkerud, V. Y. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402–420.
- Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*, 220–224.
- Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. *Developmental Science*, *8*, F1–F12.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.
- Ladefoged, P. (2006). *A course in phonetics* (5th ed.). Boston: Thomson Wadsworth.
- Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 6771–6774.
- Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*, 187–196.
- Marler, P., & Peters, S. (1977). Selective vocal learning in a sparrow. *Science*, *198*(4316), 519–521.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178.
- Moon, C., Bever, T. G., & Fifer, W. P. (1992). Canonical and non-canonical syllable discrimination by two-day-old infants. *Journal of Child Language*, *19*, 1–17.
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior & Development*, *16*, 495–500.
- Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior & Development*, *21*, 779–784.
- Nicastro, N., & Owren, M. J. (2003). Classification of domestic cat (*Felis catus*) vocalizations by naive and experienced human listeners. *Journal of Comparative Psychology*, *117*, 44–52.
- Owren, M. J., & Rendall, D. (2003). Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: Perceptual experiments with human (*Homo sapiens*) listeners. *Journal of Comparative Psychology*, *117*, 380–390.
- Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions. *Journal of the Acoustical Society of America*, *101*, 2951–2963.
- Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*(5571), 1321–1323.
- Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., et al. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 11702–11705.
- Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*, 201–236.
- Remez, R. E., & Rubin, P. E. (1993). On the intonation of sinusoidal sentences: Contour and pitch height. *Journal of the Acoustical Society of America*, *94*, 1983–1988.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–949.
- Rendall, D., Owren, M. J., & Rodman, P. S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *Journal of the Acoustical Society of America*, *103*, 602–614.
- Rendall, D., Seyfarth, R. M., Cheney, D. L., & Owren, M. J. (1999). The meaning and function of grunt variants in baboons. *Animal Behaviour*, *57*, 583–592.
- Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, *16*, 197–201.
- Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*(7079), 978–982.
- Spence, M. J., & DeCasper, A. J. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior & Development*, *10*, 133–142.

- Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*, 270–276.
- Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*, 159–164.
- Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*, 437–456.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*, 49–63.
- Zuberbuhler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology*, *113*, 33–42.