

Three-Month-Olds Prefer Speech to Other Naturally Occurring Signals

Sarah Shultz

Department of Psychology, Yale University

Athena Vouloumanos

Department of Psychology, New York University

Human infants show a preference for listening to speech, but little is known about how infants listen to other naturally occurring sounds. Here, we test infants' listening bias for speech against a range of naturally occurring sounds that share properties of speech to varying extents and we aim to better characterize the speech properties that attract infant attention. We compared 3-month-olds' listening patterns for five types of sounds: nonnative speech, rhesus macaque vocalizations, human noncommunicative vocalizations, human communicative nonspeech vocalizations, and environmental sounds. Across three experiments, 3-month-olds preferred speech to the other four types of sounds. The set of acoustic properties we measured—pitch, peak amplitude, nonzero-root mean square amplitude, frequency difference and amplitude variance—did not predict infant looking time. Our results demonstrate that young infants attend selectively to speech over many other naturally occurring stimuli, an important tool for learning language.

For human infants to learn language, they must distinguish the sounds (and gestures) that are relevant for language acquisition. Infants as young as newborns possess listening biases that direct them towards human speech. One line of research shows that infants listen longer to speech compared with a range of synthetic sounds such as filtered speech (Spence & DeCasper, 1987), warbled tones (Samples & Franklin, 1978), white noise (Butterfield & Siperstein, 1970; Colombo & Bundy, 1981), and synthetic sine-wave analogues of speech (Vouloumanos & Werker, 2004, 2007). Speech processing also elicits distinct neural responses in young infants (Dehaene-Lambertz, 2000; Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002) and newborns (Peña et al., 2003) compared to reversed speech and tone complexes. Infants thus process speech differently or preferentially compared with synthetic nonspeech sounds. Less known is whether infants prefer speech compared with other naturally occurring sounds. The only study to date examined

This research was funded by a Natural Sciences and Engineering Research Council (NSERC) of Canada grant to AV, and an NSERC USRA to SS, and was conducted in part in facilities funded by the Centre for Research in Language, Mind, and Brain.

Correspondence should be addressed to Athena Vouloumanos, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. E-mail: athena.vouloumanos@nyu.edu

3-month-olds' and neonates' preference for nonsense speech sounds that were modeled on their native language compared with rhesus macaque vocalizations (Vouloumanos, Hauser, Werker, & Martin, 2010). Three-month-olds, but not neonates, listened more to speech than to rhesus vocalizations. Here, we test infants' listening bias for speech against a range of naturally occurring sounds and aim to better characterize the factors that govern infants' early listening preference for speech. We also seek to replicate and extend our previous finding by providing the first test of a preference for nonnative (Japanese) speech sounds compared to other naturally occurring signals.

To the extent that infants prefer listening to speech compared with other sounds, their preference might be driven by any of several different properties of speech. Speech can be described on many levels, from its physical characterization as a collection of broadband frequencies that change over time to its capacity for bearing meaning and reference. In this study, we focus on a set of properties that can distinguish speech from other naturally occurring sounds: speech is a biological sound produced by an animate organism; more specifically, speech is produced by a human source; further, speech is a communicative signal, able to transfer information from a sender to a receiver; finally speech is a linguistic signal, making use of a specific repertoire of sounds produced by precise articulatory gestures. Are infant listening preferences driven by any one of these properties of speech or by a conjunction of all four? If infants' listening biases were tuned more broadly, infants might be responsive to, say, all biological signals relative to nonbiological signals. In contrast, if infant listening biases were tuned more narrowly, infants might privilege human speech over other sounds.

To test whether infants' listening biases are tuned to a broad or narrow range of sounds, we contrast speech with rhesus macaque (*Macaca mulatta*) vocalizations, human involuntary noncommunicative vocalizations, human nonspeech communicative vocalizations, and environmental sound stimuli. Each of these nonspeech sound types shares several key properties—biological (animate) origin, human source, or communicative function—with speech (see Table 1). First, calls of rhesus monkeys, though by no means representative of all nonhuman biological sounds, were selected because their similarity to speech provides a strong test for preference. Rhesus monkey vocalizations are similar to human speech in that both originate from a biological source and may be used in referential and emotionally valenced contexts (Gouzoules, Gouzoules, & Marler, 1984; Hauser, 2000). However, although rhesus calls and human speech overlap in frequencies, rhesus calls encompass a broader and higher spectral range and fewer harmonics (or overtones) than do speech sounds (Fitch & Hauser, 1995; Owren & Rendall, 2003; Rendall, Owren, & Rodman, 1998), reflecting a distinctly nonhuman source. In a previous study,

TABLE 1
Properties of Human Speech Sounds Contrasted in the Three Experiments

	<i>Speech</i>	<i>Rhesus macaque calls</i>	<i>Environmental</i>	<i>Human noncommunicative vocalizations</i>	<i>Human communicative vocalizations</i>
Linguistic	✓				
Human	✓			✓	✓
Communicative	✓	✓			✓
Biological	✓	✓		✓	✓

3-month-olds preferred English-like nonsense words to rhesus vocalizations (Vouloumanos et al., 2010). We examine whether this preference holds for nonnative speech. Second, humans often produce vocal sounds involuntarily and without intent to communicate. These sounds, such as coughs, sneezes, and yawns, are, like speech, emitted from a human source but primarily for physiological reasons and thus do not typically serve a communicative function. Third, humans also produce communicative vocalizations such as gasps of surprise, laughter, and expressions of disgust that are not typically considered speech-like. Finally, environmental sounds differ from speech in being produced by nonbiological sources such as mechanical sources (e.g., bells) or physical sources (e.g., wind, rain) that are found in infants' natural surroundings.

We attempt to minimize the familiarity of the specific speech tokens in this study by using isolated words spoken in Japanese. Japanese is a mora-timed language that is perceptually distinct in prosodic and temporal organization from the predominantly stress-timed or syllable-timed languages spoken in these infants' local community (Ramus, Nespors, & Mehler, 1999). These rhythmically defined categories of languages can be differentiated by the relative duration of vocalic and intervocalic intervals, with mora-timed languages such as Japanese being characterized by relatively equal durations of vowel and consonant segments (Ramus, 2002). More importantly, infants under six months of age are highly sensitive to the rhythmic properties of speech: they discriminate utterances from different rhythmic classes (Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998) and demonstrate a clear preference for rhythmically familiar languages (Hayashi, Tamekawa, & Kiritani, 2001; Moon, Cooper, & Fifer, 1993).

Using the five sound types—speech, rhesus calls, human noncommunicative vocalizations, human communicative vocalizations, and environmental sounds—we conducted three experiments examining listening preferences in early infancy.

EXPERIMENT 1

The first experiment examines whether infants demonstrate listening preferences when speech, rhesus calls, human noncommunicative vocalizations, and environmental sounds are contrasted using a four-way sequential preferential listening procedure. We expect that broad listening biases may lead infants to favor all three biological signals (speech, rhesus calls, and human noncommunicative vocalizations) over environmental sounds. In contrast, if infant listening biases are narrow, we expect different patterns depending on which (human source, linguistic or communicative function) is the relevant property. For example, if infants are attracted to sounds produced by humans, infants would show a relative preference for human speech and human noncommunicative vocalizations. If infants are drawn by communicative signals, they may show a relative preference for human speech and rhesus calls.

METHOD

Participants

Infants were recruited from birth lists obtained from government sources. Parents gave informed consent on behalf of their children and received a t-shirt and certificate as gifts. All participants

were healthy, full-term deliveries (40 weeks \pm 2.5 weeks), had no histories of ear infections, and had not been previously exposed to Japanese.

Twenty-three infants between the ages of 77 and 120 days ($M = 102$ d) were included in this study. An additional 15 infants were excluded from analysis due to fussiness and crying (8), parental interference (2), failure to look away from the screen to end the trial for 8 or more of the 16 test trials (2), failure to fixate on the screen for more than 1 second on two trials of the same sound type (1), failure to fixate on the screen (1), and experimenter error (1).

Materials

Four types of auditory stimuli were presented: a human speech set, a rhesus macaque vocalization set, a human noncommunicative set, and an environmental set. All sounds were resampled at 48000 Hz and equalized for mean intensity using PRAAT 4.3.12 (Boersma & Weenink, 2005). A PRAAT custom script was used to concatenate the sounds into four 40 s sound files per sound type, each consisting of 33–35 tokens separated by 600–1000 ms of silence. The selection and ordering of tokens comprising the 40 s sound files was pseudo-random: each token was selected randomly, without replacement, until all tokens had been selected. This process was repeated until enough tokens were selected to create a 40 s sound file. The same token was never played twice in a row. A summary of acoustic measures extracted from these different stimulus types is found in Table 2.

Human speech consisted of 17 tokens of Japanese words spoken by three adult female native Japanese speakers: “ho” (5 tokens), “biiru” (3), “suika” (3), “shouyu” (2), “nasu” (2), and “buta” (2). All words were spoken in adult-directed speech to ensure that infants do not show a bias towards listening to the human speech samples simply because they were infant-directed (e.g., Fernald, 1985).

Rhesus macaque vocalizations consisted of 17 tokens produced by three free-ranging adult female rhesus macaques recorded in Cayo Santiago, Puerto Rico (Marc Hauser, Harvard University): coos (5), noisy/tonal screams (3), grunts (3), arched screams (2), harmonic arches

TABLE 2
Acoustic Characteristics of Sound Stimuli Used in the Three Experiments.*

<i>Sound Type</i>	<i>Pitch (Hz)</i>	<i>Length (ms)</i>	<i>Peak</i>	<i>non-Zero RMS</i>	<i>Frequency Difference</i>	<i>Amplitude Variance</i>
Human Speech	211	418	0.389	0.258	0.0068	0.027
Rhesus Macaque	373	398	0.248	0.179	0.0063	0.051
Environmental	507	434	0.385	0.215	0.0070	0.027
Human Non-Communicative	342	395	0.419	0.199	0.0062	0.051
Human Communicative	253	347	0.400	0.289	0.0076	0.015

*For each sound set, average values are reported for: pitch, length, peak amplitude, nonzero RMS, difference (of adjacent frequency values in time), and variance (of amplitude values in time). See text for more information.

(2), and girneys (2). These calls differ from one another on valence and referential function (Hauser, 2000) and thus represent a wide range of rhesus vocalizations.

Human noncommunicative sounds consisted of 17 tokens emitted by three females: sniffs (5), coughs (3), hiccups (3), sneezes (2), yawns (2), and throat clearings (2). These sounds were selected because humans emit them primarily for physiological reasons. Exceptions such as an attention-garnering throat-clearing “ahem” presumably lie outside the grasp of infants.

Environmental sounds consisted of 17 tokens downloaded from www.findsounds.com: bells (5), breaking glass (3), running water (3), wind (2), alarms (2), and hammer sounds (2). To create the strongest test of infant preference, we included a mixed variety of environmental sounds. Sound length was shortened in SoundEdit 16 (Version 2, Macromedia, San Francisco, Calif.).

Pretest Stimulus

A pretest sound consisted of a 40 s clip of Bach’s Concerto for Violin and Orchestra No 1 in A minor (BWV 1041 - III. Allegro Assai). The purpose of the pretest was to familiarize the infants with the procedure and thus not have to eliminate the first test trial from the analysis (e.g., Cooper & Aslin, 1994; Vouloumanos & Werker, 2004).

Apparatus

Testing was conducted in a sound-attenuated 6-foot x 8-foot room. The walls were covered by curtains and the room was lit by a dim lamp situated in the front right corner of the room. Infants were seated on their caregivers’ lap 26” away from a 17–20” monitor and recorded using a digital video camera placed 6” below the monitor. Sounds were played at an average amplitude of 60dB (± 5 dB; sound pressure levels were measured using the A scale, slow setting) using two speakers placed below the monitor.

Experimenters were blind to the specific condition being tested for all infants and presented the stimuli from behind a partition while observing the infants’ behavior through a closed-circuit television. The experimenter and parents listened to masking music over Peltor HT74 headphones. An Apple G4 Macintosh computer was used to control the presentation of stimuli using Habit 7.6 (Leslie B. Cohen, University of Texas at Austin).

Design and Procedure

Infants were tested using an infant-controlled sequential preferential looking procedure (e.g., Cooper & Aslin, 1994; Vouloumanos & Werker, 2004), as successful use of this procedure has been demonstrated with infants as young as one month (Cooper & Aslin, 1994). Between every trial, a red flashing light was presented on the monitor to attract the infant’s attention. Once the infant fixated on the light a stationary black and white checkerboard was shown on the monitor simultaneously with the presentation of the sounds. Once the infant looked away from the monitor for more than two seconds or when the maximum trial length of 40 s had been reached,

the presentation of the stimuli ceased. When the infant looked back at the monitor, the presentation of the checkerboard recommenced, but this time in tandem with a new set of experimental sounds.

The entire test phase was composed of four test blocks, each of which contained one trial of each sound type presented in a random order for a total of 16 sound trials. Within one trial, sounds of a given sound type were played in predetermined pseudo-random order for a maximum of 40 s. Test blocks were presented in a semi-random order to ensure that infants never heard the same type of sound in consecutive trials.

Infants' looking time for each type of sound (human speech, rhesus macaque, human noncommunicative, or environmental) was calculated based on 30 frame-per-second, frame-by-frame coding of infant looks towards the monitor during each test trial. Missing values for a trial were replaced with the average of all infants' looking times for that particular sound trial in that particular block (e.g., the second block rhesus call trial would be replaced by the average of the remaining infants' looking time during the rhesus call trial in the second block; e.g., Johnson & Aslin, 1995; Rose & Feldman, 1987). Missing values occurred when a trial was terminated prematurely (before an infant accumulated 1 s of total looking time). Infants could have a maximum of two missing trial values and never more than one missing value per sound type. The average looking time for each of the four sound types was calculated for each infant.

RESULTS AND DISCUSSION

As a mixed ANOVA yielded no effect of sex or order (all $ps > .05$), we collapsed across these factors in subsequent analyses. A repeated measures ANOVA with sound type (i.e., speech, rhesus monkey call, human noncommunicative, and environmental) and block (1, 2, 3, 4) as within-subject factors yielded a main effect of block ($F(3, 66) = 12.67, p = .001$), indicating that infants' looking times decreased as the experiment progressed, and a main effect of sound type, $F(3, 66) = 3.22, p = .028$. A series of planned-comparison t -tests were performed to compare speech against each of the other three sound types. The comparisons revealed that infants looked longer during the presentation of the speech sounds than they did during the presentation of the human noncommunicative sounds, $t(22) = 2.8, p = .01$, and rhesus macaque calls, $t(22) = 2.4, p = .025$. The difference in looking time between speech and environmental sounds did not reach statistical significance ($t(22) = 1.26, p = .22$). There were no differences between looking time for human noncommunicative sounds, rhesus calls, and environmental sounds (all $ps > .05$). Mean looking times for each sound type are given in Table 3 and shown in Figure 1. Overall, significantly more infants showed longer looks for speech ($n = 12$) compared with environmental ($n = 5$), rhesus macaque ($n = 3$), and human noncommunicative ($n = 3$) sounds (chi-square test $\chi^2(3, N = 23) = 9.52, p = .023$) (see Figure 1).

Thus, 3-month-old infants listened longer to speech compared with rhesus macaque vocalizations and human noncommunicative sounds, suggesting that neither communicative sounds nor sounds originating from a human source were as engaging as speech for infants. With a sample of 23 infants, the difference between looking time for speech and environmental sounds did not reach statistical significance.

TABLE 3
Acoustic Measures of Each Sound Type Used in the Three Experiments and Mean Infant Looking Times for Each Sound Type*

Sound Type	Pitch (Hz)	Length (ms)	Peak	non-Zero RMS	Frequency Difference	Amplitude Variance	Infant Looking Time (s) (Expt 1)	Infant Looking Time (s) (Expt 2)	Infant Looking Time (s) (Expt 3)
Human Speech	228	408	0.249	0.175	0.00608	0.0517	14.7 (1.4)	14.8 (1.3)	16.7 (1.5)
Rhesus Macaque	373	398	0.248	0.179	0.00629	0.0512	11.6 (1.4)		
Human Non-Communicative	342	395	0.419	0.199	0.00620	0.0513	11.1 (1.4)		12.4 (1.9)
Mixed Environmental	632	442	0.371	0.221	0.00631	0.0495	12.9 (1.3)	12.7 (1.1)	
Multi-language	196	412	0.339	0.342	0.00717	0.0049		14.0 (1.3)	
Human Speech	209	435	0.579	0.256	0.00716	0.0243			
Water Environmental	362	425	0.40	0.208	0.00774	0.0040		11.3 (1.1)	
Human Communicative	253	348	0.40	0.289	0.00756	0.0155			13.0 (1.5)

* Average values are reported for pitch, length, peak amplitude, nonzero RMS, frequency difference, and amplitude variance. Infants' looking times for each stimulus set are reported for each experiment (means, with standard error in parenthesis).

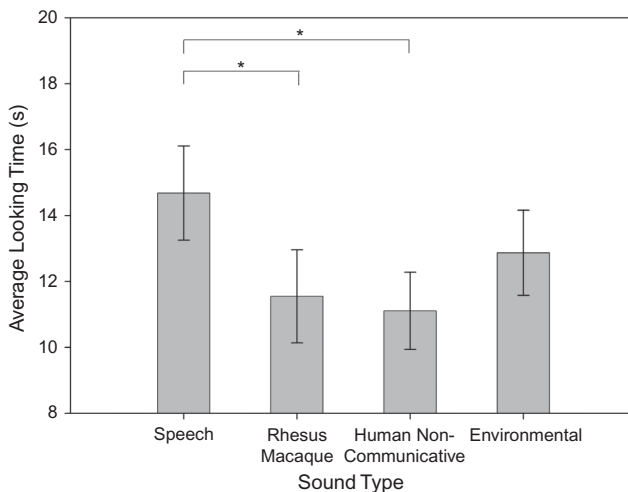


FIGURE 1 Mean looking time in seconds for each of four sound types contrasted in Experiment 1.

EXPERIMENT 2

In Experiment 2, we further examine the listening preferences of 3-month-old infants for speech and environmental sounds. To test the replicability of our results in Experiment 1, we present a new group of infants with the speech and environmental sound sets of Experiment 1. To explore listening preferences for a more diverse corpus of speech and environmental sounds we present infants with two additional sound sets: (1) a set of environmental sounds that are generated by water (e.g., breaking waves, dripping rain water) and (2) speech sounds spoken in several nonnative languages (Japanese, Italian and Russian).

METHOD

Participants

Infants were recruited as for Experiment 1. All participants had not been exposed to Japanese, Italian, or Russian. Twenty-five infants between the ages of 95 and 133 d ($M = 116$ d) were included in this study. An additional 19 infants were excluded from analysis due to fussiness or crying (7), failure to end the trial before the maximum trial length had been reached on 8 or more of the 16 test trials (7), failure to fixate on the screen for more than 1 s on two trials of the same sound type (2), experimenter error (2), and parental interference (1).

Materials

Single-language speech sounds were those of Experiment 1.

Multilanguage speech consisted of words spoken by three adult female speakers whose native language was Russian, Japanese, or Italian. Two sets of multilanguage human speech were used, but because there were no differences in looking time across the two sets, they were analyzed together. The first set (A) consisted of 17 tokens of the words “juu” (5), “te” (3), “poka” (3), “viso” (2), “nasu” (2), and “sreda” (2). The second set (B) consisted of 17 single tokens of the words “nuca,” “re,” “ceci,” “te,” “viso,” “kuma,” “shougo,” “juu,” “mame,” “nasu,” “neko,” “ni,” “svet,” “poka,” “sreda,” “doch,” and “pol.”

Water environmental sounds consisted of 17 tokens originating from a water source: dripping water (5), lake water (3), boiling water (3), stream water (2), water being poured (2), and splashing water (2) sounds.

Mixed environmental sounds were those of Experiment 1.

Pretest stimulus was identical to that used in Experiment 1.

Sounds were edited for length, adjusted for intensity and sampling rate, and concatenated into 40 s files as in Experiment 1. Acoustic measures are reported in Table 2.

Apparatus

See Experiment 1.

Design and Procedure

Infants were tested using an infant-controlled sequential preferential looking procedure as in Experiment 1. Test trials, trial order, and test blocks were created according to the parameters of Experiment 1 such that every infant heard 16 trials, four of each sound type, presented across four test blocks in a pseudo-random order so as not to hear the same sounds in consecutive trials.

Infants' looking time for each type of sound (mixed environmental, water environmental, multi-language speech, and single-language speech) was calculated based on 30 frame-per-second, frame-by-frame coding of infant looks towards the monitor during each test trial. Coded looking time data were not acquired for three infants due to a technical malfunction that resulted in unreadable videotapes. Given the high correlation ($r = .97, p = .001$) of looking times generated by the online Habit software and offline frame-by-frame coding, online looking times were included in the analysis for these three infants. As in Experiment 1, missing values for a trial were replaced by taking the average of all infants' looking times for that particular sound trial in that block. Infants had a maximum of two missing trial values and never more than one missing value per sound type.

RESULTS AND DISCUSSION

As there were no significant effects of sex, order, or multi-language version A or B (all $ps > .05$), we collapsed across these factors in subsequent analyses. A repeated measures ANOVA on sound type and block revealed a significant main effect of block ($F(3,72) = 20.47, p < .001$), reflecting a decrease in overall looking in later blocks and a significant main effect of sound type ($F(3,72) = 2.78, p = .047$). Planned comparisons revealed that infants looked longer during the presentation of multilanguage speech ($t(24) = 2.08, p = .049$) and single-language speech

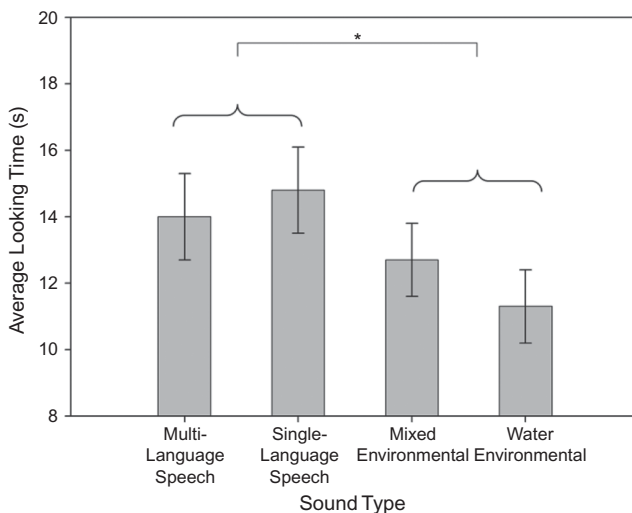


FIGURE 2 Mean looking time in seconds for each sound type in Experiment 2.

($t(24) = 3.10$, $p = .005$) than they did during the presentation of the water environmental sounds. As in Experiment 1, with a sample of 25 infants, no looking time differences were observed between multi-language speech, single-language speech, and mixed environmental sounds trials (all $ps > .05$). Mean looking times for each sound type are reported in Table 3 and shown in Figure 2.

As a final probe, we collapsed across sounds that were part of the same overarching category to test for infant looking time differences between speech (multi- and single-language) and environmental sounds (mixed and water environmental). A paired samples t-test revealed that infants looked longer during the presentation of speech sounds than to environmental sounds, $t(24) = 2.57$, $p = .017$ (see Table 3 and Figure 2).

Given that the single-language speech and mixed environmental sound trials were identical across Experiments 1 and 2, we combined the looking time data from Experiments 1 and 2 ($n = 48$). A repeated measures ANOVA yielded a main effect of sound ($F(1, 47) = 4.24$, $p = .045$), with longer average looking times for single-language speech sounds than for mixed environmental sounds. The numerical difference in looking time between single-language speech and mixed environmental sounds observed across both individual experiments (Experiments 1 and 2) reached significance when the samples were combined.

EXPERIMENT 3

In Experiment 3, we examine whether speech is a preferred stimulus relative to other types of human vocalizations. Specifically, we compare speech with human communicative vocalizations including laughter, sounds of agreement, and surprise. These sounds are prosodically rich and available in the child's early environment (Papoušek, Papoušek, & Haekel, 1987). We include the human noncommunicative vocalizations as a third human stimulus set.

METHOD

Participants

Infants were recruited as for Experiments 1 and 2. Fourteen infants between the ages of 106 and 120 days old ($M = 111$ d) were included in this study. An additional seven infants were excluded from analysis due to crying or fussing (3), failure to end the trial before the maximum trial length had been reached on 6 or more of the 12 test trials (3), and experimenter error (1).

Materials

Human speech sounds were those of Experiment 1.

Human noncommunicative vocalizations were those of Experiment 1.

Human communicative vocalizations consisted of 17 vocalizations spoken by three adult female speakers: laughter (5), agreement (3), inquiring (3), surprise (2), disagreement (2), and disgust (2).

Pretest stimulus was identical to that used in Experiment 1.

Sounds were edited for length, adjusted for intensity and sampling rate, and concatenated into 40 s files as in Experiment 1. Acoustic measures are reported in Table 2.

Apparatus

See Experiment 1.

Design and Procedure

Infants were tested using an infant-controlled sequential preferential looking procedure as in Experiment 1. Test trials, trial order, and test blocks were created according to the parameters of Experiment 1 such that every infant heard 12 trials, four of each sound type, presented across four test blocks in a pseudo-random order so as not to hear the same sounds in consecutive trials.

Infants' looking time for each type of sound (speech, human noncommunicative and human communicative vocalizations) was calculated based on 30 frame-per-second, frame-by-frame coding of infant looks towards the monitor during each test trial. As in Experiment 1, missing values for a trial were replaced by taking the average of all infants' looking times for that particular sound trial in that block. Infants had a maximum of two missing trial values and never more than one missing value per sound type.

RESULTS AND DISCUSSION

A repeated measures ANOVA on sound type and block revealed a significant effect of block ($F(3, 39) = 4.19, p = .012$), reflecting a decrease in looking overall in later blocks, and a significant effect of sound type ($F(2, 26) = 4.87, p = .016$). Planned comparisons revealed reliably longer looking during human speech trials than either human noncommunicative

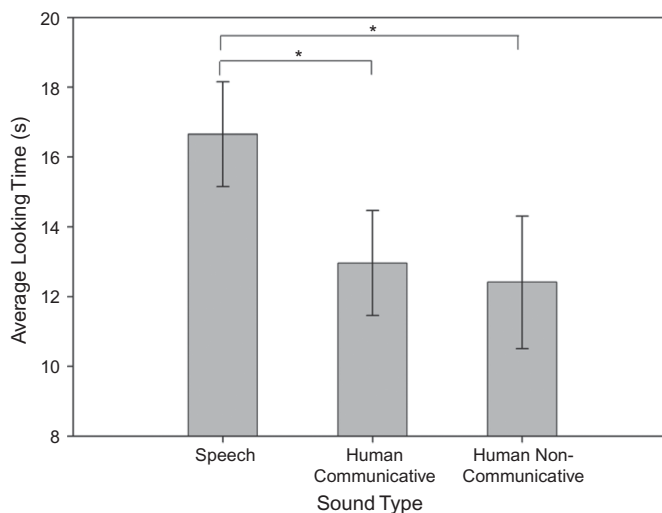


FIGURE 3 Mean looking time in seconds for each sound type in Experiment 3.

($t(13) = 2.66, p = .020$) or communicative vocalizations ($t(13) = 2.88, p = .013$). Mean looking times for each sound type are reported in Table 3 and shown in Figure 3. Finally, of the 14 infants tested, more infants listened longer to the speech stimuli (9) as compared to either the noncommunicative vocalizations (2) or the communicative vocalizations (3) (significant by a multinomial test, $p = .003$). Three-month-old infants listened longer to speech compared with both human communicative and human noncommunicative sounds, suggesting that even communicative sounds originating from a human source were not as engaging as speech for infants.

ACOUSTIC ANALYSES

We examined the extent to which differences in acoustic measures account for infants' looking time patterns for the different sound types presented in all three experiments. Using MitSyn software (Digivox, Montreal, Canada), we extracted several measures from the sounds: pitch which correlates with fundamental frequency, the nonzero root mean square and peak amplitude which reflect perceived loudness, a difference score reflecting the frequency differences between two adjacent points in time which is an index of the complexity of the sound (a smaller number means a smooth sound with low complexity, a larger number a rougher sound with higher complexity), variance which reflects the amplitude variation across time about the mean, and stimulus length (values are reported in Table 3).

Because pitch can influence listening times (e.g., Fernald & Kuhl, 1987), we ensured there were no differences in the mean pitch between the mixed environmental sounds and water environmental sounds using a student's t -test, $t(24) = 1.95, p > .05$. As such, we collapsed across both environmental sound types in subsequent acoustic analyses. A series of one-way factorial analyses of variance (ANOVAs) revealed significant differences between sound sets in pitch

($F(4,131) = 11.65, p < .001$), in peak intensity ($F(4,131) = 2.50, p < .05$), in frequency difference score ($F(4,131) = 3.45, p = .01$), and in amplitude variance ($F(4,131) = 13.05, p < .001$), but not in nonzero root mean square or length ($ps > .05$). After Bonferroni correction for multiple comparisons, the pitch of environmental sounds was higher than both speech ($p < .001$) and communicative vocalizations ($p < .001$). Rhesus pitch was also higher than speech ($p = .039$). Differences in peak intensity did not survive Bonferroni correction for multiple comparisons. After Bonferroni correction, the frequency difference score of the communicative sounds was significantly higher than that of rhesus calls ($p = .041$) and of noncommunicative sounds ($p = .021$). Rhesus calls had higher amplitude variance than speech ($p < .001$), environmental sounds ($p < .001$), and communicative sounds ($p < .001$). Human noncommunicative sounds also had higher variance than speech ($p < .001$), environmental sounds ($p < .001$), and communicative sounds ($p < .001$).

To explore whether infant listening time was sensitive to any of these measures, we examined the degree to which each of these acoustic measures (pitch, root mean square, peak amplitude, frequency difference score and variance) and length correlated with infants' looking time across all sound types across the three experiments. The only significant correlation was a negative relationship between length of tokens and infant looking time ($r = -.103, p = .001$). In particular, there was a significant negative correlation between token length and looking time for environmental sounds ($r = -.14, p < .05$) and speech sounds ($r = -.187, p < .01$). We found no significant correlations between any of the acoustic measures and infant looking time (see Table 3). We conclude that the differences in infant looking time observed for the different sound types are not modulated by the acoustic factors that we examined.

GENERAL DISCUSSION

Across three experiments, 3-month-old infants listened longer to nonnative speech sounds compared with a range of other sounds, including rhesus monkey vocalizations, human noncommunicative vocalizations, and human nonspeech communicative vocalizations. Infants also preferred speech to environmental sounds, although this effect was modulated by the kinds of environmental sounds presented. By three months, speech is a preferred stimulus for infants, eliciting their interest over many other natural nonspeech sounds.

These results extend and refine our understanding of the listening biases at work in early infancy. Previous studies with infants around three months of age had demonstrated a preference for speech compared with white noise (Colombo & Bundy, 1981) and with melodic nonspeech analogues (Vouloumanos & Werker, 2007). A recent study found a preference for native speech sounds over rhesus monkey calls in 3-month-olds but not in newborns (Vouloumanos et al., 2010). The current results demonstrate that the bias for speech extends to nonnative sounds (Japanese, a language unfamiliar to the infants tested in this study) and holds even in comparison to other human-produced sounds, such as laughter and yawns.

Although infants showed an overall preference for speech compared to environmental sounds, this difference seemed less pronounced relative to the other comparisons. Why should sounds like breaking glass, running water, or bells be engaging? Infants may have failed to perceive the mixed environmental sounds as a coherent category of sounds due to the sounds having been produced naturally (wind and rain), mechanically (breaking glass), and electronically (alarms

and bells) (Ballas & Howard, 1987; Shafiro, 2004). The large range of possible environmental sounds might result in “little reason to expect that a common set of features exists” (Ballas & Howard, 1987, p. 102). Whereas the human-produced and rhesus vocalizations share features such as frequency range and relatively sustained fundamental frequency, such shared features were absent from the mixed environmental sounds.

Three month olds’ preference for speech may be a product of their accumulated experience with specific properties of their native languages, English and French. At the same time, other studies show that infants do not recognize the specific sounds, sound sequences, or stress of their native language until 6–10 months (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Friederici, Wessels, & Svenkerud, 1993; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 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 for Japanese speech in these relatively younger 3-month-olds in our studies derived directly from properties of the specific languages to which they were exposed. Instead infants may become attuned to general properties of spoken language shared across languages.

Our findings also shed light on the properties of sounds that attract infant attention in early life. The different nonspeech sounds we contrasted in this study shared some properties of speech such as biological or human source and communicative function to different extents (see Table 1). Neither of these properties alone or in combination was as interesting to infants as speech. Even the human communicative sounds in the final experiment, which share human, biological, and communicative properties with speech, were not as interesting to infants as speech.

The acoustic properties that underlie infant listening preferences are not yet clear. We examined a set of acoustic properties: pitch, perceived loudness, sound complexity and amplitude variation; none of these measures correlated with infant listening time. This suggests that infant looking times are modulated either by a different acoustic property that we have not examined or by several properties of sounds in combination. For example, global alternations between rapid frequency transitions and steady state frequencies that characterize consonant-vowel alternations typical of human speech might draw infant interest (Ladefoged, 2006). It will be essential in future work to systematically explore which acoustic features of these complex signals draw infants’ attention.

The discovery of a preference for speech versus other naturally occurring stimuli suggests several important questions. First, it is of interest to determine what kinds of listening preferences are present from birth (e.g., Vouloumanos & Werker, 2007) to examine whether listening specificity arises after exposure to speech in the first few months of life, suggesting a perceptual narrowing as has been observed in other domains (Hespos & Spelke, 2004; Pascalis, de Haan, & Nelson, 2002; Werker & Tees, 1984). In recent studies, neonates preferred speech to synthetic nonspeech sounds (Vouloumanos & Werker, 2007) but showed equal interest in speech and rhesus vocalizations, consistent with somewhat broader initial preferences (Vouloumanos et al., 2010).

Another important question for future research is whether infants’ preference for speech is robust outside the laboratory in the natural environment, which is composed of multiple competing sounds. Moreover, in the natural environment sounds are perceived in combination with signals from other modalities (e.g., vision, touch) (Fowler & Dekle, 1991; Rosenblum, 2008).

Pairing naturally occurring sounds with objects in the environment (e.g., faces) could potentially enhance or alter the preferences we observed here.

Despite these remaining questions, a bias for attending to speech versus other naturally occurring stimuli is an important tool for learning language. Atypical speech listening preferences have been linked to impairments in language development in children with autism spectrum disorders (Kuhl, Coffey-Corina, Padden, & Dawson, 2005; Paul, Chawarska, Fowler, Cicchetti, & Volkmar, 2007). By extracting and attending to species-specific linguistic signals from their auditory environments, infants may gain more exposure to speech signals and thus analyze them more completely, ultimately perceiving finer gradations in the speech stimulus (Gibson, 1979). Early proclivities towards the processing of human speech sounds could underlie our capacity to process and produce spoken language.

ACKNOWLEDGMENTS

We thank Pierre Ahad for acoustic analyses, Kris Onishi and Gary Marcus for helpful discussion, Marc Hauser for providing the rhesus monkey vocalizations, Sarah Racine for recording the human communicative vocalizations, Hannah Schreier for coding infant looking time data for Experiment 1, and the members of the McGill Infant Development Centre and the NYU Infant Cognition and Communication Lab. Finally, we are especially grateful to the participating parents and infants.

REFERENCES

- Ballas, J. A., & Howard, J. H., Jr. (1987). Interpreting the language of environmental sounds. *Environment and Behavior*, 19, 91–114.
- Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer* (4.3.12 ed.) [Computer software].
- 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.
- Colombo, J. A., & Bundy, R. S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior & Development*, 4(2), 219–223.
- Cooper, R. P., & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, 65(6), 1663–1677.
- Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience*, 12(3), 449–460.
- Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298(5600), 2013–2015.
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development*, 8(2), 181–195.
- Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development*, 10(3), 279–293.
- Fitch, W., & Hauser, M. D. (1995). Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on “honest” advertisement. *American Journal of Primatology*, 37(3), 191–219.
- Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3), 816–828.
- Gelfand, H., & Vouloumanos, A. (2009, October). *Conversations with parrots: The effect of the vocalizing source on speech perception in infants and adults*. San Antonio, TX: Cognitive Development Society.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

- Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signaling in the recruitment of agonistic aid. *Animal Behavior*, *32*(1), 182–193.
- Hauser, M. D. (2000). A primate dictionary? Decoding the function and meaning of another species' vocalizations. *Cognitive Science*, *24*(3), 445–475.
- Hayashi, A., Tamekawa, Y., & Kiritani, S. (2001). Developmental change in auditory preferences for speech stimuli in Japanese infants. *Journal of Speech and Hearing Research*, *44*(6), 1189–1200.
- Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature*, *430*(6998), 453–456.
- Johnson, S. P., & Aslin, R. N. (1995). Perception of object unity in 2-month-old infants. *Developmental Psychology*, *31*(5), 739–745.
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 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*(3), 402–420.
- 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*(1), 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.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178.
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior & Development*, *16*(4), 495–500.
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766.
- 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*(4), 380–390.
- Papoušek, M., Papoušek, H., & Haekel, M. (1987). Didactic adjustments in fathers' and mothers' speech to their 3-month-old infants. *Journal of Psycholinguistic Research*, *16*(5), 491–514.
- 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.
- Paul, R., Chawarska, K., Fowler, C., Cicchetti, D., & Volkmar, F. (2007). "Listen my children and you shall hear": Auditory preferences in toddlers with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1350–1364.
- Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (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*(20), 11702–11705.
- Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In *Proceedings of Speech Prosody 2002* (pp. 115–120). April 11–13, 2002, Aix-en-Provence, France.
- Ramus, F., Nespore, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292.
- 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*(1), 602–614.
- Rose, S. A., & Feldman, J. F. (1987). Infant visual attention: Stability of individual differences from 6 to 8 months. *Developmental Psychology*, *23*(4), 490–498.
- Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*(6), 405–409.
- Samples, J. M., & Franklin, B. (1978). Behavioral responses in 7 to 9 month old infants to speech and non-speech stimuli. *Journal of Auditory Research*, *18*(2), 115–123.
- Shafiro, V. (2004). Perceiving the sources of environmental sounds with a varying number of spectral channels. Unpublished doctoral dissertation, City University of New York.
- 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*(2), 133–142.

- Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development, 81*, 517–527.
- Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science, 7*(3), 270–276.
- Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science, 10*(2), 159–164.
- 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*(1), 49–63.