Rayna H. Friendly[1]
Drew Rendall[2]
Laurel J. Trainor[1,3]

[1]*Department of Psychology*
Neuroscience and Behaviour
McMaster University
1280 Main Street West, Hamilton,
Ontario, Canada, L8S 4L8
E-mail: ljt@mcmaster.ca

[2]*Department of Psychology*
University of Lethbridge
4401 University Drive, Lethbridge,
Alberta, Canada, T1K 3M4

[3]*Rotman Research Institute*
Baycrest Centre
3560 Bathurst Street, Toronto, Ontario,
Canada, M6A 2E1

Learning to Differentiate Individuals by Their Voices: Infants' Individuation of Native- and Foreign-Species Voices

ABSTRACT: *The ability to discriminate and identify people by their voice is important for social interaction in humans. In early development, learning to discriminate important differences in a number of socially relevant stimuli, such as phonemes and faces, has been shown to follow a common pattern of experience-driven perceptual narrowing, where the discrimination of native stimuli improves, while the discrimination of foreign stimuli worsens. The aim of the present study was to investigate whether similar perceptual narrowing occurs for discriminating individuals by voice. We tested the ability of English-speaking adults and English-learning 6- and 12-month-olds to discriminate either native-species (human) or foreign-species (rhesus monkey [Macaca mulatta]) individuals by their vocalizations. Between 6 and 12 months of age, the ability to discriminate monkey voices decreased significantly and there was a non-significant trend for improved ability to discriminate human voices. The results support the hypothesis of experience-driven perceptual narrowing for voice individuation during the first year after birth.* *© 2013 Wiley Periodicals, Inc.* Dev Psychobiol

Keywords: *voice discrimination; perceptual narrowing; infant; adult; primate; development; plasticity; cross-species; experience; rhesus monkey; human*

INTRODUCTION

The ability to identify individual people quickly and effortlessly is important for human social interaction and likely enhanced survival in our ancestors. One way that humans distinguish individuals is by the unique sound of each person's voice. This ability to distinguish people by voice is especially important in situations where the talker is not clearly visible to the listener, including the example from recent times of talking on the telephone. Because voice recognition plays an important role in social interaction, understanding the developmental trajectory for voice discrimination is important for understanding social and communicative development. Much evidence indicates that complex perceptual processing, such as that involved in face recognition (Kelly et al., 2007; Pascalis, de Haan, & Nelson, 2002; Simpson, Varga, Frick, & Fragaszy, 2010), speech sound categorization (Kuhl et al., 2006; Werker & Tees, 2005), musical pitch and rhythmic processing (Hannon & Trehub, 2005a,b; Lynch & Eilers, 1992; Lynch, Eilers, Oller, & Urbano, 1990; see Hannon & Trainor, 2007; Trainor & Corrigall, 2010; Trainor & Unrau, 2012 for reviews), intersensory (face/voice) processing (Lewkowicz & Ghazanfar, 2006; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009), and action perception (Loucks & Sommerville, 2012) is strongly influenced during the first year after birth by experience with the specific sounds and objects in the environment in which the infant is developing. Specifically, the ability to make perceptual discriminations relevant to the particular

---

Manuscript Received: 2 February 2013
Manuscript Accepted: 3 August 2013
Correspondence to: Laurel Trainor
Article first published online in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/dev.21164 • © 2013 Wiley Periodicals, Inc.

language, musical system and facial features experienced in the native environment improves, while the ability to make discriminations relevant to foreign (and not to native) languages, musical systems, and faces diminishes (see Lewkowicz & Ghazanfar, 2009; Scott, Pascalis, & Nelson, 2007, for reviews). This process is known as perceptual narrowing (Lewkowicz & Ghazanfar, 2009).

There is extensive evidence for perceptual narrowing for face processing. Adults demonstrate a bias for better discrimination of own-race and own-species faces compared to the discrimination of other-race and other-species faces (e.g., Meissner & Brigham, 2001; Pascalis & Bachevalier, 1998). Findings from behavioral (e.g., Bar-Haim, Ziv, & Hodes, 2006; Hayden, Bhratt, Joseph, & Tanaka, 2007; Kelly et al., 2005, 2007, 2009; Pascalis et al., 2002; Sangrigoli & De Schonen, 2004; Simpson et al., 2010) and ERP (e.g., de Haan, Pascalis, & Johnson, 2002; Scott, Shannon, & Nelson, 2006; Scott & Monesson, 2010) studies with infants suggest mixed evidence for such biases at 3 months, clearly developing biases by 6 months, and largely adult-like biases by 9 months of age. For example, when familiarized with one face, and then subsequently presented with the familiarized and a novel face shown from different vantage points (ranging from facing toward the left to the right) both from the same race, Kelly et al. (2007, 2009) found that at 3 months of age Caucasian and Chinese infants can discriminate individuals from another race as accurately as individuals from their own race. However, Caucasian 6-month-olds discriminate both Caucasian and Chinese faces but fail the same tests with African or Middle Eastern faces, and Chinese 6-month-olds discriminate both Chinese and Caucasian but not African faces (they were not tested on the Middle Eastern faces). By 9 months of age, both Caucasian and Chinese infants show no evidence of discrimination for other-race faces. A similar developmental progression to the Kelly et al. studies has also been reported for own-species compared to other-species faces. In this case, 6-month-old infants have been reported to be equally good at discriminating forward-facing faces of individuals from native (human) and foreign (monkey) species but show no evidence of discrimination for monkey faces by 9 months of age (Pascalis et al., 2002). In addition, the ability to detect whether the auditory and visual aspects of vocalizing dynamic faces are synchronous is species- and language-general at 6 months of age but by 8–11 months, there is evidence of this intersensory ability for native but not for foreign species and languages (Lewkowicz & Ghazanfar, 2006; Pons et al., 2009).

Despite the importance of discriminating and identifying people by voice, little is known about how this ability develops. However, several studies indicate that infants recognize familiar voices in their environment. By 36–39 weeks gestational age, the fetus can discriminate its mother's voice from that of another female (Kisilevsky et al., 2003) and can discriminate between a male and a female voice (Lecanuet, Granier-Deferre, Jacquet, Capponi, & Ledru, 1993). Newborn infants recognize and prefer their mother's voice to that of a female stranger when the voices are producing infant-directed speech (DeCasper & Fifer, 1980; Mehler, Bertoncini, Barriere, & Jassik-Gerschenfeld, 1978; Spence & Freeman, 1996). Although newborns show no evidence of preferring sentences spoken by their father compared to those spoken by an unfamiliar male, they can tell these voices apart (DeCasper & Prescott, 1984).

When listening to voices, infants are learning to extract two different kinds of information, one related to learning the language spoken and the other to identifying people by their voice. As far as learning language, infants show early sensitivity to prosodic features, preferring to listen to their native language over languages with different prosodic structures (Mehler et al., 1988; Moon, Panneton Cooper, & Fifer, 1993). Furthermore, many studies show that learning phonemic categories follows a trajectory of perceptual narrowing similar to that of face processing. Adults show well-developed specialization for native phonemic categories in that they have difficulty discriminating between two speech sounds that fall into different phonemic categories in a foreign language but within a single phonemic category in their native language (Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Kuhl, 1998; Kuhl et al., 2006; Werker & Lalonde, 1988; Werker et al., 2007; Werker & Tees, 1984, 2005). For consonant contrasts, there is no evidence for perceptual narrowing until after 6 months of age, but narrowing is established by 10–12 months of age (for reviews see Curtin & Werker, 2007; Kuhl, 2008). For vowel contrasts, narrowing appears to occur somewhat earlier, beginning around 4 months of age (Kuhl, 2004; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994). Interestingly, perceptual narrowing also occurs for sign language (Palmer, Fais, Golinkoff, & Werker, 2012). English-hearing infants can discriminate American Sign Language (ASL) hand shape distinctions at 4 months, but fail the comparable test at 14 months, of age, while their ASL-learning counterparts can discriminate the distinctions at both ages (Palmer et al., 2012).

Despite the rich literature on language learning, only one study to our knowledge has addressed whether perceptual narrowing occurs for voice identification. Johnson, Westrek, Nazzi, and Cutler (2011) examined the ability of 7-month-old infants to discriminate two different unfamiliar talkers of the same sex. They demonstrated that Dutch 7-month-olds showed a native-language processing bias, discriminating between two unfamiliar female voices speaking Dutch sentences, but failed the same test with two unfamiliar female voices speaking Japanese or Italian sentences. Although further studies are needed with younger and older infants in order to map out the developmental trajectory, this study suggests that, just as infants acquire an own-race bias for discriminating faces, they acquire an own-language bias for discriminating voices.

In the present paper, we examine the related question of whether infants also acquire an own-species bias for the ability to discriminate individuals by voice, similarly to how they acquire an own-species bias for discriminating faces. In particular, we compared the discrimination of human vocalizations to those of a phylogenetically close foreign species, the rhesus monkey. Importantly, there is behavioral (Vouloumanos, Hauser, Werker, & Martin, 2010) and neural (Minagawa-Kawai et al., 2011) evidence that young infants can discriminate human and non-human primate vocalizations. Unlike in the case of better discrimination of voices in a familiar versus foreign language (Johnson et al., 2011), where basic characteristics of the human vocal apparatus remain the same across languages spoken, perceptual narrowing for human compared to rhesus monkey voice discrimination would indicate specialization for processing characteristics of sounds made by the adult human vocal tract that are unique to our species (such as a relatively larger oral cavity and descended larynx, see review by Fitch, 2000). We hypothesized that, if the development of voice discrimination occurs through a process of perceptual narrowing, then ability to discriminate native-species voices would improve, while the ability to discriminate foreign-species voices would worsen between 6 and 12 months of age. This age range was chosen because previous research has found evidence of narrowing over this period in other auditory domains, such as perceiving consonant phonemes and musical rhythms (e.g., Curtin & Werker, 2007; Hannon & Trehub, 2005a,b).

METHODS

Participants

Twenty-four adults between 18 and 40 years of age ($M = 21$ years, $SD = 6.26$ years; 12 females), each reporting normal hearing and providing informed consent, participated in this experiment. Participants were drawn from first- and second-year undergraduate psychology classes and received course credit for their participation. All reported (Canadian) English as their only spoken and understood language.

The final infant sample contained 48 infants between 5.5 and 6.5 months of age (M = 6.01 months, $SD = .24$ months; 25 females), and 48 infants between 11.5 and 12.5 months of age (M = 11.98 months, $SD = .17$ months; 19 females), all reported to have normal hearing. Infants were randomly assigned to listen to either human or rhesus monkey voices. All infants were reported as hearing English spoken in their home environment 98–100% of the time. In the sample of 12-month-olds, three additional infants were tested, but excluded from the final sample due to fussiness ($n = 1$) or failure to pass training ($n = 2$), all three from the human voice condition. In the sample of 6-month-olds, 11 additional infants were tested, but excluded from the final sample due to fussiness ($n = 3$, human condition; $n = 1$, monkey condition) or failure to pass training ($n = 4$, human; $n = 3$, monkey). All procedures were approved by the McMaster Research Ethics Board. Informed consent was obtained from parents of infant participants.

Stimuli and Apparatus

Human Voice Stimuli. Voice recordings were made of eight monolingual English-speaking Canadian female adults using a Neutrik AKG (Emotion D 770) microphone, and Felt Tip Sound Studio 2.1 software via a USB Audio/Midi interface (US—122; sampling rate = 44.1 kHz; resolution = 16-bit) on a Macintosh computer (Power Mac G5, OS X version 10.3.9). Female human voices were chosen over male voices as the monkey voice stimuli that we had access to were also of female voices. Each female speaker produced the word *balloon* with six different intonational contours, obtained by reciting the following six sentences: 1. This is a */balloon/*. 2. Do you want a */balloon/*? 3. What a great */balloon/*! 4. Have you seen a */balloon/* today? 5. This */balloon/* is very light! 6. */Balloon/* stands sell lots of balloons! The tokens of the word *balloon* were extracted from the six sentences using Cool Edit Pro and were normalized for intensity. The same word was used for all human vocalizations so that linguistic information would not be informative in any way. The intonational differences were created to ensure that infants were not using intonation as a cue to individuate the voices. The word *balloon* was chosen as the human voice stimulus because it is an infant-friendly word that can be spoken easily with different intonation contours in a variety of carrier sentences.

After acoustic analyses using Praat software (Boersma & Weenink, 2009), two pairs of female voices (pairs 1 and 2) were chosen for use as test stimuli such that the set of six /balloon/ tokens from each of the two voices within a voice pair were matched for mean duration ($M = .694$ s, $SD = .010$ s; $M = .511$ s, $SD = .067$ s, for the two pairs, respectively) and minimum F0 ($M = 164.04$ Hz, $SD = 27.86$ Hz; $M = 183.36$ Hz, $SD = 19.14$ Hz), maximum F0 ($M = 348.61$ Hz, $SD = 78.18$ Hz; $M = 279.95$ Hz, $SD = 39.23$ Hz) and mean F0 ($M = 234.57$ Hz, $SD = 35.84$ Hz; $M = 221.64$ Hz, $SD = 14.50$ Hz). In other words, the average of the set of tokens from voice pair 1 was matched to the average of the set of tokens from voice pair 2, for example. This matching was important to ensure that participants would need to use attributes other than these characteristics to discriminate the voices.

Four conditions (1A, 1B, 2A, 2B) were constructed for testing infants. For each voice pair (1, 2), there were two conditions (A, B) such that for condition A one speaker's voice in the pair served as the "background" voice, while the other served as the "change" voice, and vice versa for condition B (see Procedure Section).

Rhesus Monkey Voice Stimuli. Rhesus monkey (Macaca mulatta) voices were chosen as the foreign-species test stimuli for the following reasons. First, the ability of non-human primates to perceive formants (an acoustic feature of voices) is similar to that of humans (Ghazanfar et al., 2007; Owren, 1990; Sommers, Moody, Prosen, & Stebbins, 1992). Second, there is evidence of vocal recognition by rhesus monkeys of both individuals and kin (Rendall, Rodman, & Emond, 1996). Third, six or more vocal samples of the same call category ("coo") from each of six female rhesus monkeys were available from author DR (for methodology on obtaining these recordings see Owren & Rendall, 2003; Rendall et al., 1996). Using these rhesus monkey stimuli, Owren and Rendall (2003) compared human adults' ability to discriminate rhesus coos and human vowel sounds, as well other rhesus vocalizations, such as screams. They showed that human adults found the rhesus coos more difficult to discriminate than human vowels, but easier to discriminate than rhesus screams (Owren & Rendall, 2003).

Example spectrograms comparing the human and rhesus monkey stimuli used in the present study are shown in Figure 1. Both voice types contain a stable fundamental frequency, F0 (the bottom-most band of energy in the spectrograms), above which are harmonic overtones. For both voice types, the amplitudes of the harmonics are filtered by the unique resonances, or formants, of each individual's vocal tract shape and length. The "coo" calls were edited using Cool Edit Pro (sampling rate = 44.1 kHz, resolution = 16-bit) and normalized for intensity. Acoustic analyses using Praat led to the choice of two pairs of primate voices (six different tokens of the coo call for each individual monkey). Each pair was matched, as with the human voices, on mean duration ($M = .302$ s, $SD = .047$ s; $M = .274$ s, $SD = .061$ s, for the two pairs, respectively) and minimum F0 ($M = 282.36$ Hz, $SD = 66.28$ Hz; $M = 502.48$ Hz, $SD = 18.93$ Hz), maximum F0 ($M = 351.84$ Hz, $SD = 49.87$ Hz; $M = 562.59$ Hz, $SD = 32.49$ Hz) and mean F0 ($M = 320.50$ Hz, $SD = 48.72$ Hz; $M = 542.04$ Hz, $SD = 26.30$ Hz) F0. Four conditions (3A, 3B, 4A, and 4B) were constructed for testing infants, as with the human voices, such that for each voice pair (3 and 4), one voice in the pair served as the "background" and the other as the "change" voice for condition A, and vice versa for condition B.

In the adult procedure, the voice samples were presented using the program Presentation on an LG (Antec, Windows XP, Fremont, CA) computer through headphones (Sennheiser Electronic GmBH & Co.KG, Wedemark, Germany). In the infant procedure, the voice stimuli were presented on a Macintosh computer (Power Mac G4) and the experiment was run using in-house software. The computer was connected to a NAD Stereo Integrated Amplifier (C352) and a single Westsun loudspeaker, and a button box that was used by the experimenter for signaling infant responses to the computer. The button box controller and loudspeaker were located in a sound-attenuating chamber (Industrial Acoustics Company, Inc.), with the loudspeaker located on top of a

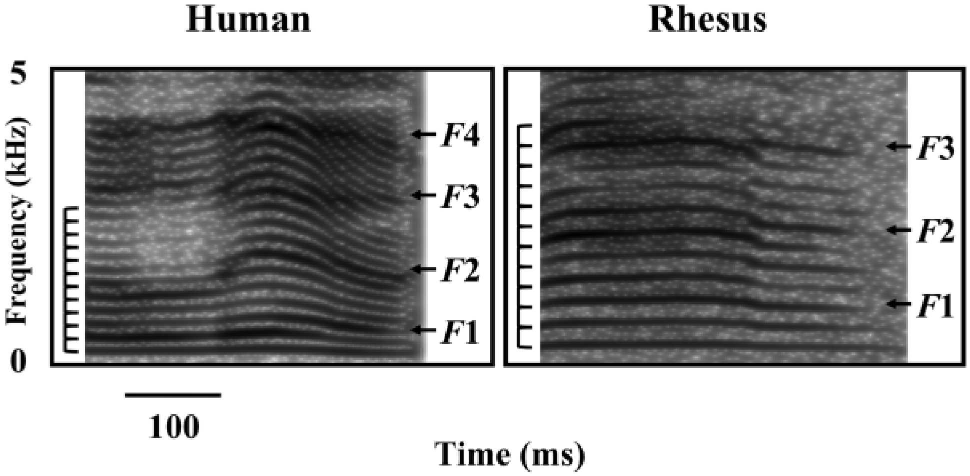

FIGURE 1 Spectrograms of the word "balloon" spoken by an adult human female and a "coo" vocalization produced by an adult female rhesus monkey. The sounds are generally similar in showing a stable fundamental frequency (F0), the lowest band on the spectrogram, with rich harmonic overtones that are filtered by the resonances, or formants, of the vocal tract (F1–F4). The sounds are different in that the F0 is higher in the rhesus monkey sounds (375 vs. 230 Hz in the human female sounds), which yields more widely spaced harmonic overtones in the rhesus monkey sounds (marked by the vertical grids at the left of each spectrogram). The formant frequencies are also higher in the rhesus monkey sounds. Note that the human vocalizations used were about 600 ms whereas the rhesus monkey vocalizations were about 300 ms.

Plexiglas-covered cabinet containing animated toys that served as reinforcement for infant behavioral responses. The computer controlled the animated toys as well as the lights that made them visible (see Procedure Section for more details).

Procedure

Adult Same-Different Task. Adults were tested using a within-subjects design that measured their discrimination of both the human and monkey vocalizations. Each adult completed two blocks of 40 same-different trials, one containing human and the other rhesus monkey voices. The blocks were presented in random order across participants. On the 20 "same" trials, two voice tokens were presented from the same individual and on the 20 "different" trials, two tokens were presented from two different individuals. On each trial, the two voices presented were always from a voice pair that was matched on duration and mean, minimum and maximum F0 (see Stimuli), and tokens were chosen randomly with the constraints that no token pair be presented more than once, no pair from a "same" trial contain the same token repeated twice, and that equal numbers of each token per voice be presented during testing. Participants were told that they would hear two voice tokens and that they were to press one button if the voices belonged to the same individual, or a second button if the voices belonged to two different individuals. For data analysis, for each adult, percent correct scores in the same-different paradigm were converted to d' separately for the human and monkey voice conditions.

Infant Conditioned Head Turn Task. Because infants are not able to perform the task used with adults, infants were tested in a conditioned head turn paradigm (Werker et al., 1998), using a between-subjects design where infants were tested on their discrimination of human or monkey vocalizations. After the experimenter obtained informed consent from parents, infants were randomly assigned to one of eight stimulus conditions, counter-balancing species (human, rhesus), voice pair (1, 2 for human; 3, 4 for rhesus) and which voice was the background and which the change voice within the pair (A, B). Thus, at each age, 6 infants were tested in each of the 8 conditions.

During the testing phase, the six /balloon/ or /coo/ tokens from the background voice were repeated in random order with a stimulus onset asynchrony (SOA) of 1750 ms through the loudspeaker located 90° to the infant's left. The infant was seated on his or her parent's lap across from and facing the experimenter. To eliminate parental or experimenter influence on the infant's behavior, both parent and experimenter listened to masking music over headphones during the procedure. The background tokens were played continuously throughout the experiment. Once the infant was attentive and facing forward (toward the experimenter), the experimenter pressed a button, indicating to the computer that the infant was ready for a trial. Of the 24 trials, half (12) were change trials, on which the background voice was replaced by one of the six tokens of the change voice for one repetition. Each of the six tokens of the change voice was presented twice during the testing phase in random order. The other half of the trials (12) were control (no-change) trials, on which the background voice continued such that control trials were indistinguishable from the background. Change and control trials were presented in a quasi-random order with the constraint that no more than two control trials be presented in a row. The experimenter pressed a second button every time the infant turned their head at least 45 degrees to the left. Head turns on change trials (i.e., hits) within 1 s of the onset of the change-voice were rewarded with 2 s of an animated toy and light display. However, head turns occurring on control trials (i.e., false alarms) were recorded but not rewarded by the computer. For data analysis, the number of hits (head turns during change trials) and false alarms (head turns during control trials) were converted into d' scores.

Prior to the testing phase, infants completed a training phase designed to familiarize them with being rewarded with an animated toy display for turning their head when the voice changed from one individual to another (one human to another when the infant was in the human voice condition, and one monkey to another when the infant was in the monkey voice condition). The training phase did not contain control trials, and only two of the six tokens from the change voice were used. In addition, the change voice was 8 dB louder than the repeating background voice, such that this noticeable difference would attract the infant's attention towards the loudspeaker. Infants were required to make four correct head-turn responses in a row within 20 training trials in order to proceed to the testing phase, where all six tokens of each voice were used and change and background tokens were presented at equal intensities. Infants who did not pass this training criterion were excluded from the final data set (see Participants Section).

RESULTS

Adults

A paired-sample t-test revealed that adults' d' scores for human voices were significantly larger than for rhesus monkey voices, $t(23) = 17.90$, $p < .001$ (human, $M = 2.76$, $SE = .13$; rhesus, $M = .37$, $SE = .10$), suggesting that they discriminated between human voices much more easily than between rhesus monkey voices (see Fig. 2). One-sample t tests revealed that adult's d' scores were significantly above chance levels (expected d' value of 0) for both voice types (both P's $< .002$). Thus, although adults were able to successfully discriminate between voices for both human and rhesus monkey voice conditions, their performance was much better for voices from their native species compared to voices from the foreign species.

Infants

One-sample t-tests revealed that d' scores were significantly above chance levels ($p < .002$) for each voice

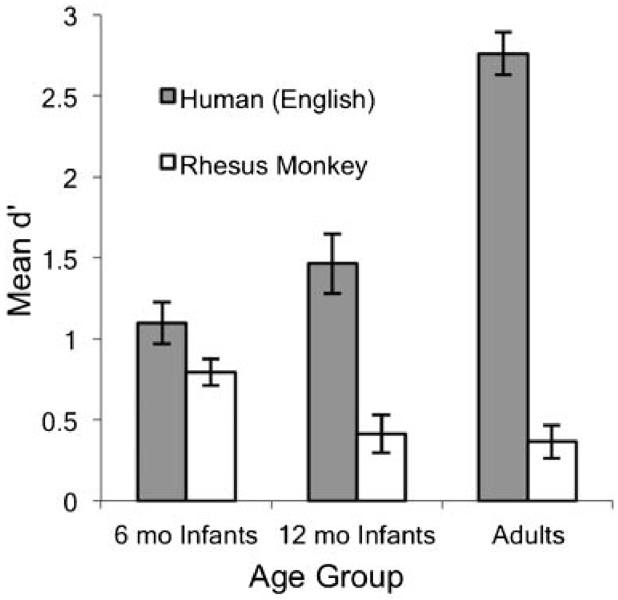

FIGURE 2 Adult ($n = 24$) and infant (6- and 12-month-olds, $n = 48$ each) discrimination of rhesus monkey and human voices, as measured by a same-different task and a conditioned head turn task, respectively. Although it is not appropriate to compare discrimination across voice types (see text), discrimination of human voices improved with age whereas discrimination of monkey voices declined with age. Error bars represent within-subjects SEM for adults and between-subjects SEM for infants.

type for both the 6- and 12-month-olds after Bonferroni correction ($\alpha = .0125$). An analysis of variance (ANOVA) with d' score as the dependent variable and age (6 and 12 months) and voice type (human, rhesus monkey) as independent variables revealed only a significant interaction between age and voice type, $F(1, 92) = 7.95$, $p = .006$, $\eta^2 = .08$ (see Fig. 2). Follow up independent sample t tests conducted separately for each voice type revealed that the d' scores for rhesus monkey voices were significantly lower at 12 than at 6 months of age, $t(46) = 2.69$, $p = .01$ (6 months, $M = .79$, $SE = .08$; 12 months, $M = .41$, $SE = .12$), suggesting a decrease in discrimination ability. There was also a trend for d' scores for human voices to increase between 6 and 12 months of age, although this effect did not reach significance $t(46) = -1.63$, $p = .11$ (6 months, $M = 1.10$, $SE = .13$; 12 months, $M = 1.46$, $SE = .18$). The results indicate that infants' ability to discriminate between two foreign-species voices decreases between 6 and 12 months of age, supporting the hypothesis of experience-driven perceptual narrowing for voice discrimination during the first year after birth.

DISCUSSION

This is the first study to show that infants' ability to discriminate individuals by voice becomes increasingly specialized for human voices between 6 and 12 months of age. Specifically, between 6 and 12 months of age we found a significant decline in the ability to individuate rhesus monkeys by their vocalizations, as well as a non-significant trend for increased ability to discriminate humans by their vocalizations. The finding that perceptual narrowing occurs by the end of the first year for voice discrimination roughly parallels findings for face discrimination (Kelly et al., 2007; Pascalis et al., 2002; Simpson et al., 2010), intersensory (face/voice) matching (Lewkowicz & Ghazanfar, 2006; Pons et al., 2009), phonemic categorization (Kuhl et al., 2006; Werker & Tees, 2005), and musical pitch and rhythm processing (Hannon & Trehub, 2005a,b; Lynch et al., 1990; for reviews see Hannon & Trainor, 2007; Trainor & Corrigall, 2010; Trainor & Unrau, 2012). Furthermore, it complements a previous study indicating that 7-month-old infants demonstrate an advantage for discriminating voices in their native compared to a foreign language (Johnson et al., 2011), by showing that perceptual narrowing also applies at a species level, with infant perception becoming specialized for processing unique characteristics of human voices during the first year after birth.

Although the ability to discriminate the monkey voices decreased significantly between 6 and 12 months of age, 12-month-olds remained above chance levels for discriminating the monkey voices. This is in contrast to Pascalis et al. (2002) who found that 6-month-olds, but not 9-month-olds and adults, showed evidence of discriminating between two monkey faces above chance levels. This seeming discrepancy might be explained by methodological differences. Pascalis et al. (2002) used a visual paired comparison (VPC) procedure whereas a conditioned head turn was used in the present study. In Pascalis et al.'s study, the VPC involved habituating an infant to one face for a fixed period of 20 s (stopping and starting the count whenever the infant looked away; adults only received 5 s of familiarization), and then comparing their looking time to either the same face or a novel face. The Conditioned Head-Turn (CHT) procedure has been reported to be better able than habituation-based methods to separate infant boredom (or disinterest) from perceptual difficulty (Werker, Polka, & Pegg, 1997) and is thus likely a more robust measure of infants' perceptual discrimination than the VPC procedure. The CHT procedure also enables researchers to collect data from multiple trials (e.g., 24 trials) as opposed to the 1 or 2 trials in a VPC procedure. Furthermore, although

Pascalis et al.'s (2002) study found no evidence that adults could discriminate monkey faces, Mondlock, Maurer, and Ahola (2006) showed that, if tested with more sensitive methodologies, adults can discriminate between monkey faces, albeit more poorly than between human faces. Specifically, they used a task in which participants were required to indicate whether two faces appearing one at a time belonged to the same individual or two different individuals. Therefore, it is possible that older infants and adults might have some ability to discriminate monkey faces, but that this ability was not revealed using the VPC procedure. Furthermore, amount of familiarization time has also been shown to influence the discrimination abilities of older infants. Fair, Flom, Jones, and Martin (2012) found that although 12-month-olds failed to demonstrate discrimination of unfamiliar monkey faces following a 20 s familiarization period, when familiarization was increased to 40 s, 12-month-olds were successfully able to discriminate the monkey faces.

In the present study, we examined changes across age in infants' ability to discriminate human and rhesus monkey voices, but we did not directly compare performance on human versus monkey voices. This comparison was not appropriate as it is very difficult to equate the discriminability of two voice types. Equating discriminability of voices is difficult for two main reasons. First, it is difficult to determine whether human and monkey voice pairs are equally discriminable in general in the absence of experience with voices from either species. In fact, it is possible that human voices are intrinsically easier to process than rhesus monkey voices. Although there are fundamental similarities between rhesus monkey coos and human vowel sounds (Owren & Rendall, 2003, also see Stimuli), considerable differences exist between the vocal anatomies of these species and, therefore, the acoustic properties of their vocalizations, suggesting that voice quality differences could be easier to detect in human than rhesus voices. For example, human vocal folds are longer and thicker than those in rhesus monkeys, permitting a greater degree of variation between individuals' voices (Schon Ybarra, 1995; Titze, 1994). Human vowel sounds also typically contain two to four times the number of harmonics compared to rhesus coo calls (see Fig. 1). These extra harmonics enable human vowels to be better defined than rhesus coos, potentially increasing the saliency of the variation between individuals in detailed aspects of their particular resonant patterns (Owren & Rendall, 2003).

The second difficulty in equating human and rhesus voices is more specific to our stimuli. Human vocalizations are typically more complex and contain consonant and vowel sounds whereas rhesus vocalizations contain only vowel-like sounds. We chose to use realistic vocalizations for each species. Our human stimuli consisted of the word "balloon" that contains both consonant and vowel sounds, whereas the monkey stimuli consisted of the very common rhesus coo call that only contains vowel-like sounds. Indeed, the possibility that some acoustic differences may be intrinsically more difficult to discriminate than others is consistent with the finding that acoustic salience influences perceptual narrowing, for example, during the processing of nasal-place phonetic distinctions (Narayan, Werker, & Beddor, 2010). Our human vocalizations were also longer, on average, than the monkey vocalizations, so they might have contained more identifying information (about 600 vs. 300 ms, respectively). In order to determine if length of utterance and the presence of consonants influences comparisons across species, and if our results would generalize to other human vocalizations, future studies could compare narrowing for human voices uttering sounds with vowels only to those uttering sounds with both consonants and vowels.

Nevertheless, the present study revealed a significant decrease in d' scores between 6 and 12 months for monkey voices, but not for human voices, so acoustic differences between the human and monkey voices cannot explain the present finding of perceptual narrowing for monkey voice discrimination between 6 and 12 months of age. This question could be investigated further in a cross-over design in which perceptual narrowing was investigated in both human and rhesus monkey species. If experience is the predominant variable rather than intrinsic properties of human and monkey voices, it would be expected that rhesus monkeys would improve at discriminating monkey voices, but get worse at discriminating the human voices, with increasing age.

We cannot ascertain from the present study whether the timeline of perceptual narrowing for voice discrimination might be underway earlier than that for processing consonant contrasts, faces from foreign species, musical scales and musical meters, for which there is no evidence of narrowing until after 6 months of age (e.g., Curtin & Werker, 2007; Hannon & Trehub, 2005a,b; Hannon & Trainor, 2007; Pascalis et al., 2002). Indeed, examination of Figure 1 appears to indicate that our data more closely resemble the timeline found for the perceptual narrowing of faces from other races, vowel contrasts and the lexical tones within tonal languages such as Mandarin, for which narrowing appears to be underway prior to 6 months after birth (e.g., Kelly et al., 2007, 2009; Kuhl, 2004; Kuhl et al., 1992; Polka & Werker, 1994; Yeung, Chen, & Werker, 2013). It is therefore important to test discrimination of these stimuli by infants younger than 6 months of age. Unfortunately, the head turn procedure used in the present study is not suitable for testing auditory discrimination in younger infants (Werker et al., 1998), but future studies could be conducted utilizing appropriate methods to determine whether narrowing might be underway for voice discrimination prior to 6 months of age.

A further question of interest for future research is when the sensitive period ends for specialization for human voice discrimination and the influence of experience on this period of sensitivity. It is possible that some plasticity remains throughout the lifespan, such that even in adults, intensive experience with voices from another species would lead to better discrimination of individual voices within that species, although pre-narrowing levels might not be attainable. There is evidence that plasticity remains in infancy beyond the point at which perceptual narrowing appears to be accomplished. For example, Pascalis et al. (2005) demonstrated that the loss of ability to discriminate between primate faces observed between 6 and 9 months of age could be prevented by 3 months of exposure during this period to primate faces. Also, Anzures et al. (2012) showed that it was possible to reinstate the ability of 8- to 10-month-olds to recognize foreign-race faces with 3 weeks of exposure to these faces. In the language domain, 1 month of interaction with Mandarin Chinese speakers reversed the decline in ability to distinguish between Mandarin phonemes in English-learning infants previously observed at 12 months (Kuhl, Tsao, & Liu, 2003). Similarly, 2 weeks of exposure to foreign musical rhythms at 12 months reinstated sensitivity to these rhythms whereas a similar amount of exposure had no effect in adulthood (Hannon & Trehub, 2005b).

One of the first studies to directly investigate the effect of environmental exposure on sensitive periods for faces in a design that strictly controlled experience was conducted by Sugita (2008). In this study, infant monkeys were reared with no exposure to live faces for periods ranging from 6 to 24 months of age. Prior to exposure, the monkeys discriminated photographs of human faces as easily as photographs of monkey faces. After the end of the period of total deprivation for face stimuli, 1 month of subsequent exposure to either human or monkey faces resulted in the monkeys only being able to discriminate the face type to which they were exposed for that month. This demonstrates that the period of plasticity for face specialization can be extended for at least 2 years in the complete absence of experience. These results, in addition to the aforementioned studies by Pascalis et al. (2005), Anzures et al. (2012) and Kuhl et al. (2003), suggest that the timing of exposure as well as the type and amount of exposure affect perceptual narrowing. These variables should also be explored with respect to voices to determine the generality of the effects of timing, type, and amount of exposure on sensitive periods.

Finally, questions remain as to whether the results of the present study generalize to foreign human languages, familiar voices, and to male in addition to female voices. It would also be interesting to know whether young infants can discriminate non-primate voices that are less similar to human voices than rhesus monkey voices, and whether the ability to discriminate voices from different species follows the same developmental trajectory as for primate voices.

CONCLUSIONS

The results indicate that young infants have the important social ability to discriminate individuals by voice, and that it is shaped by the auditory environment. Specifically, we found a significant decrease in the ability to discriminate foreign-species (rhesus monkey) voices between 6 and 12 months of age, accompanied by a non-significant trend for the ability to discriminate native-species (human) voices to increase.

NOTES

This research was supported by a grant to L.J.T. from the Natural Sciences and Engineering Research Council of Canada. We thank Lauren Kutcher and Andrea Unrau for assistance in data collection.

REFERENCES

Anzures, G., Wheeler, A., Quinn, P. C., Pascalis, O., Slater, A. M., Heron-Delaney, M., … Lee, K. (2012). Brief daily exposures to Asian females reverses perceptual narrowing for Asian faces in Caucasian infants. Journal of Experimental Child Psychology, 112, 484–495.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. Psychological Science, 17, 159–163.

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. Infant Behavior and Development, 18(3), 339–350. doi: 10.1016/0163-6383(95)90022-5

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer Version 5.1.04. Computer program. Available from http://www.praat.org/

Curtin S., & Werker, J. F. (2007). The perceptual foundations of phonological development. In M. G. Gaskell (Ed.), The oxford handbook of psycholinguistics (pp. 579–599). Oxford, UK: Oxford University Press.

de Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. Journal of Cognitive Neuroscience, 14(2), 199–209.

DeCasper A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. Science, 208(4448), 1174–1176.

DeCasper A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. Developmental Psychobiology, 17(5), 481–491.

Fair, J., Flom, R., Jones, J., & Martin, J. (2012). Perceptual learning: 12-month-olds' discrimination of monkey faces. Child Development, 83(6), 1996–2006.

Fitch, W. T. (2000). The evolution of speech: A comparative review. Trends in Cognitive Sciences, 4(7), 258–267.

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., & Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. Current Biology, 17, 425–430.

Hannon E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. Trends in Cognitive Sciences, 11(11), 466–472.

Hannon E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. Psychological Science, 16(1), 48–55.

Hannon E. E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults. Proceedings of the National Academy of Sciences of the United States of America, 102(35), 12639–12643.

Hayden, A., Bhatt, R. S., Joseph, J. E., & Tanaka, J. W. (2007). The other-race effect in infancy: Evidence using a morphing technique. Infancy, 12, 95–104.

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. Developmental Science, 14(5), 1002–1011.

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? Journal of Experimental Child Psychology, 104(1), 105–114.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy. Psychological Science, 18(12), 1084–1089.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., … Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. Developmental Science, 8, F31–F36.

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., … Wang, Z. (2003). Effects of experience on fetal voice recognition. Psychological Science, 14(3), 220–224.

Kuhl, P. K. (1998). Effects of language experience on speech perception. The Journal of the Acoustical Society of America, 103(5), 2931.

Kuhl, P. K. (2008). Linking infant speech perception to language acquisition: Phonetic learning predicts language growth. In P. McCardle, J. Colombo, & L. Freund (Eds.), Infant pathways to language: Methods, models, and research directions (pp. 213–2243). New York: Erlbaum.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. Nature Reviews Neuroscience, 5, 831–843.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Developmental Science, 9(2), F13–F21.

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. Proceedings of the National Academy of Sciences, 100(15), 9096–9101.

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.

Lecanuet, J. P., Granier-Deferre, C., Jacquet, A. Y., Capponi, I., & Ledru, L. (1993). Prenatal discrimination of a male and a female voice uttering the same sentences. Early Development and Parenting, 2(4), 217–228.

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(17), 6771–6774.

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. Trends in Cognitive Sciences, 13(11), 470–478.

Loucks J., & Sommerville, J. A. (2012). Developmental changes in the discrimination of dynamic human actions in infancy. Developmental Science, 15(1), 123–130.

Lynch M. P., & Eilers, R. E. (1992). A study of perceptual development for musical tuning. Attention, Perception and Psychophysics, 52(6), 599–608.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. Psychological Science, 1(4), 272–276.

Mehler, J., Bertoncini, J., Barriere, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. Perception, 7(5), 491–497.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. Cognition, 29, 144–178.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. Psychology, Public Policy and Law, 7(1), 3–5.

Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical brain imaging reveals general auditory and language-specific processing in early infant development. Cerebral Cortex, 21(2), 254–261.

Mondlock, C. J., Maurer, D., & Ahola, S. (2006). Becoming a face expert. Psychological Science, 17(11), 930–934.

Moon, C., Panneton Cooper, R., & Fifer, W. P. (1993). Two-day-olds prefer their native language. Infant Behavior and Development, 16, 495–500.

Narayan, C., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. Developmental Science, 13(3), 407–420.

Owren, M. J. (1990). Acoustic classification of alarm calls by vervet monkeys (cercopithecus aethiops) and humans (homo sapiens): II. Synthetic calls. Journal of Comparative Psychology, 104(1), 29–40.

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.

Palmer, S. B., Fais, L., Golinkoff, R. M., & Werker, J. F. (2012). Perceptual narrowing of linguistic sign occurs in the 1st year of life. Child Development, 83(2), 543–553.

Pascalis, O., & Bachevalier, J. (1998). Face recognition in primates: A cross-species study. Behavioural Processes, 43, 87–96.

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.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. Proceedings of the National Academy of Sciences of the United States of America, 102(14), 5297–5300.

Polka L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20(2), 421–435.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. Proceedings of the National Academy of Sciences of the United States of America, 106(26), 10598–10602.

Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kinin free-ranging rhesus monkeys. Animal Behavior, 51(5), 1007–1015.

Sangrigoli S., & De Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. Journal of Child Psychology and Psychiatry, 45(7), 1219–1227. doi: 10.1111/j.14697610.2004.00319

Schon Ybarra, M. (1995). A comparative approach to the non-human primate vocal tract: Implications for sound production. In E. Zimmermann, J. D. Newman, & U. Jürgens (Eds.), Current topics in primate vocal communication (pp. 185–198). New York: Plenum.

Scott, L. S., & Monesson, A. (2010). Experience-dependent neural specialization during infancy. Neuropsychologia, 48, 1857–1861.

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(4), 197–201.

Scott, L. S., Shannon, R. W., & Nelson, C. A. (2006). Neural correlates of human and monkey face processing by 9-month-old infants. Infancy, 10(2), 171–186.

Simpson, E. A., Varga, K., Frick, J. E., & Fragaszy, D. (2010). Infants experience perceptual narrowing for nonprimate faces. Infancy, 16(3), 318–328.

Sommers, M. S., Moody, D. B., Prosen, C. A., & Stebbins, W. C. (1992). Formant frequency discrimination by Japanese macaques (macaca fuscata). Journal of the Acoustic Society of America, 91(6), 3499–3510.

Spence, M. J., & Freeman, M. S. (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. Infant Behavior and Development, 19(2), 199–212.

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. Proceedings of the National Academy of Sciences, 105(1), 394–398.

Titze, I. R. (1994). Principles of voice production. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Trainor, L. J., & Corrigall, K. A. (2010). Music acquisition and effects of musical experience. In M. Riess-Jones & R. R. Fay (Eds.), Springer handbook of auditory research: Music perception (pp. 89–128). Heidelberg: Springer.

Trainor L. J., & Unrau, A. J. (2012). Development of pitch and music perception. In L. Werner, R. R. Fay, & A. N. Popper (Eds.), Springer handbook of auditory research: Human auditory development (pp. 223–254). New York: Springer.

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. Child Development, 81(2), 517–527.

Werker J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. Developmental Psychology, 24(5), 672–683.

Werker, J. F., Polka, L., & Pegg, J. E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. Early Development and Parenting, 6(34), 171–178.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. Cognition, 103(1), 147–162.

Werker, J. F., Shi, R., Desjardins, R., Pegg, J. E., Polka, L., & Patterson, M. (1998). Three methods for testing infant speech perception. In A. M. Slater (Ed.), Perceptual development: Visual, auditory, and speech perception in infancy (pp. 389–420). London: UCL Press.

Werker J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 79(1), 49–63.

Werker J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. Developmental Psychobiology, 46(3), 233–251.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. Journal of Memory and Language, 68, 123–139.