Supporting Information

Sliwa et al. 10.1073/pnas.1008169108

SI Materials and Methods

Text S1: Comparison of the Time Spent Looking at the Face Matching the Voice Relative to the Other Face. Data were analyzed with MATLAB custom-written scripts and the Statistics Toolbox (MathWorks). Here, we analyzed the data without taking into account preferences that monkeys displayed for individual faces. Therefore, we assessed whether the proportion of time spent on matching face was above chance level (50%). To this end, we performed a t test on data for all trials for all monkeys pooled together. To compare the number of saccades and duration of the longest fixation on the matching picture compared with the nonmatching picture, a paired-sign test was performed. Saccades were defined as increases in gaze velocity >30°/s with minimum amplitude of 1°. Fixations were defined as the gaze position in the interval between two saccades.

Monkeys looked at the matching face significantly longer compared with the other face (match, 658 ms, 51.8%; nonmatch, 612 ms, 48.2%). Although this difference is relatively small, it differed significantly from chance [one-sample t test, $t(1{,}726) = 2.31$, $P = 0.02$]. Moreover, the number of saccades and the duration of the longest fixation were higher on the matching face than on the nonmatching face (respectively $P = 0.014$ and $P = 0.045$, sign tests).

Text S2: Is the Matching Effect Driven by Gender Characteristics or Identity Characteristics? The limited amount of individuals we could use as stimuli in our experiment forced us to present male and female monkeys and humans. Here we tested whether the perceived face–voice matching effect was driven by gender matching or by identity matching. To do so, we separated trials comparing same gender faces from trials comparing different gender faces. Then, as in the main analysis, we tested the modulation of face preferences by the identity of the voice, for each face of each pair.

This analysis revealed that monkeys spend more time looking at a face when preceded by its voice in same gender trials [$t(49) = 3.54$, $P = 0.0009$] as well as in different gender trials [$t(53) = 2.94$, $P = 0.005$]. When analyzing separately trials with monkey stimuli from trials with human stimuli, it appears that results are similar to the joint results: in same gender monkey [$t(21) = 2.14$, $P = 0.04$] and human [$t(27) = 2.96$, $P = 0.006$] trials and in different gender monkey [$t(9) = 2.49$, $P = 0.03$] and human [$t(43) = 2.23$, $P = 0.031$] trials.

Text S3: Training Paradigm. Animals first learned to complete exploration trials with abstract audio and visual stimuli. Visual training stimuli were presented in pairs subtending 10° × 24°. The virtual exploration window around the pair was 12° × 26° of visual angle. It was composed of the two pictures (in upper and lower parts) surrounded by ≈4° black surround and a central 12° × 2° black area between the two pictures considered as belonging neither to one nor to the other. Auditory training and test stimuli were presented at an intensity of 55–65 dB (A-weighted) sound pressure level (SPL) at the subject's ear, measured with a Brüel and Kjær 2239A Integrating Sound Level Meter (www.bksv.com) from two speakers located 56 cm in front of the subject and symmetrically 45 cm apart.

The goal of the training task was to maintain the animal's gaze on the middle of the screen during the audio playback and allow it to freely explore two images presented together on the screen after the audio playback. The subject was to start by directing its gaze to a small white square presented on the center of the screen. Then it was required to fixate (±1°–4°) while a 2-s audio sample was played. After a delay of 0.2 s, two pictures were presented vertically and equally spaced from the center (Fig. S2). The subject could freely explore these pictures during 1.5 s, as long as its gaze was maintained within the boundaries of a virtual window around the pair of pictures corresponding to the black area in Fig. S2. In a pilot study, we tested conditions in which we did not reward keeping gaze inside of the virtual window around the pair of pictures. In this case, the monkey completed fewer trials than we needed to perform statistical tests regarding the richness of our stimuli sets. Thus, reward was given to ensure monkeys' attention to the computer monitor where visual stimuli were displayed and not to provide reinforcement for a particular exploration pattern. This training paradigm was pursued until the monkeys could readily complete 150 trials.

Text S4: Eye Tracking Calibration Procedure. Each session started by a calibration of signal offset and gain using an analog amplifier with modulating gain and offset connected in series to the ISCAN and REX Software. To do so, the monkey was rewarded for fixating on a small spot (±1°–4°) appearing at the center of the screen (offset calibration) and then equally distributed 15° in four directions: left (abscissa gain), right, up (ordinate gain), and down. Eye movements were not filtered for the analyses.

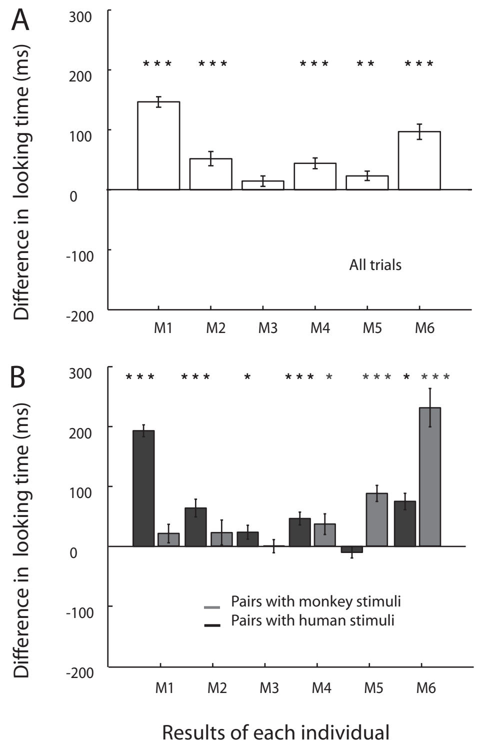

Fig. S1. Results per individual (M1–M6) for matching the identity of the voice with the identity of the face. Error bars represent the SEM (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). (*A*) Mean looking time difference between the congruent condition and the noncongruent condition for all of the pairs (n_{1-5} = 18 pairs, n_6 = 14 pairs). The abscissa indicates chance expectation. (*B*) The same as *A* but separately analyzed for pairs with monkey (n_{1-5} = 6 pairs, n_6 = 2 pairs, light shading) and human (n_{1-6} = 12 pairs, dark shading) stimuli.

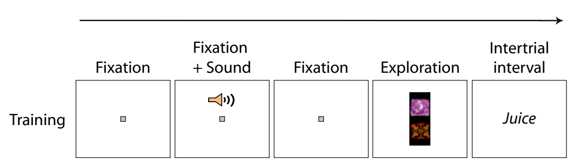

Fig. S2. Training sessions. Subjects learned to initiate a trial by fixating on a white square on the screen and then continuing to fixate for the 2-s sound duration and finally freely exploring two pictures for 1.5 s within the virtual window represented by the black area.

Table S1. Face preferences

Monkey	ANOVA on preferences for some individual faces over others	Significant confidence intervals
M1	$F_1 = 2.36$, df = 6, $P = 0.029$	$CI_{0.95}(\mu_{1-3}) = [0.65; 23.77]$
M2	$F_2 = 4.65$, df = 6, $P = 0.0001$	$CI_{0.95}(\mu_{1-3}) = [2.62; 32.75]$
		$CI_{0.95}(\mu_{2-3}) = [2.63; 32.03]$
		$CI_{0.95}(\mu_{4-7}) = [8.22; 46.47]$
		$CI_{0.95}(\mu_{5-7}) = [4.63; 41.21]$
		$CI_{0.95}(\mu_{6-7}) = [2; 38.95]$
M3	$F_3 = 4.3$, df = 6, $P = 0.0003$	$CI_{0.95}(\mu_{1-3}) = [7.38; 28.6]$
		$CI_{0.95}(\mu_{4-7}) = [0.98; 28.27]$
M4	$F_4 = 1.34$, df = 6, $P = 0.24$	None
M5	$F_5 = 15.65$, df = 6, $P = 0$	$CI_{0.95}(\mu_{1-3}) = [5.57; 20.15]$
		$CI_{0.95}(\mu_{2-3}) = [4.82; 19.37]$
		$CI_{0.95}(\mu_{4-5}) = [18.55; 37.17]$
		$CI_{0.95}(\mu_{4-6}) = [10.37; 28.96]$
		$CI_{0.95}(\mu_{4-7}) = [7.49; 25.99]$
		$CI_{0.95}(\mu_{5-7}) = [1.91; 20.33]$
M6	$F_6 = 5.34$, df = 5, $P = 9.10^{-5}$	$CI_{0.95}(\mu_{1-2}) = [14.53; 45.36]$
		$CI_{0.95}(\mu_{4-6}) = [3.53; 33.25]$
		$CI_{0.95}(\mu_{4-7}) = [0.51; 30.83]$

Table S2. Change over time

Time course bins	Overall	Monkey stimuli	Human stimuli
0–300	$t(103) = 1.62$, $P = 0.054$	$t(31) = 0.8$, $P = 0.21$	$t(71) = 1.42$, $P = 0.08$
100–400	$t(103) = 1.33$, $P = 0.094$	$t(31) = 0.7$, $P = 0.24$	$t(71) = 1.15$, $P = 0.13$
200–500	$t(103) = 0.13$, $P = 0.45$	$t(31) = 0.003$, $P = 0.50$	$t(71) = 0.13$, $P = 0.46$
300–600	$t(103) = 0.13$, $P = 0.45$	$t(31) = 0.3$, $P = 0.36$	$t(71) = -0.01$, $P = 0.5$
400–700	$t(103) = 1.16$, $P = 0.12$	$t(31) = 0.8$, $P = 0.21$	$t(71) = 0.89$, $P = 0.19$
500–800	$t(103) = 3.3$, $P = 0.0005$	$t(31) = 1.6$, $P = 0.053$	$t(71) = 2.91$, $P = 0.002$
600–900	$t(103) = 4.77$, $P = 3 \times 10^{-6}$	$t(31) = 2.2$, $P = 0.016$	$t(71) = 4.21$, $P = 4 \times 10^{-5}$
700–1,000	$t(103) = 5.46$, $P = 2 \times 10^{-7}$	$t(31) = 2.8$, $P = 0.004$	$t(71) = 4.67$, $P = 6 \times 10^{-6}$
800–1,100	$t(103) = 5.29$, $P = 3 \times 10^{-7}$	$t(31) = 3.3$, $P = 0.0009$	$t(71) = 4.25$, $P = 3 \times 10^{-5}$
900–1,200	$t(103) = 4.78$, $P = 3 \times 10^{-6}$	$t(31) = 3.0$, $P = 0.002$	$t(71) = 3.81$, $P = 0.0001$
1,000–1,300	$t(103) = 3.82$, $P = 0.0001$	$t(31) = 3.1$, $P = 0.002$	$t(71) = 2.87$, $P = 0.003$
1,100–1,400	$t(103) = 3$, $P = 0.002$	$t(31) = 2.9$, $P = 0.003$	$t(71) = 2.14$, $P = 0.018$
1,200–1,500	$t(103) = 2.85$, $P = 0.003$	$t(31) = 2.6$, $P = 0.006$	$t(71) = 2.05$, $P = 0.022$

Table S3. Individual specificities

Voice–face match	M1	M2	M3	M4	M5	M6
Overall	$t_1(17) = 16.9$, $P = 2 \times 10^{-12}$	$t_2(17) = 4.29$, $P = 0.0002$	$t_3(17) = 1.59$, $P = 0.065$	$t_4(17) = 4.78$, $P = 8 \times 10^{-5}$	$t_5(17) = 2.91$, $P = 0.005$	$t_6(13) = 7.63$, $P = 2 \times 10^{-6}$
Monkey stimuli	$t_1(5) = 1.33$, $P = 0.12$	$t_2(5) = 1.07$, $P = 0.17$	$t_3(5) = 0.032$, $P = 0.48$	$t_4(5) = 2.09$, $P = 0.045$	$t_5(5) = 6.49$, $P = 0.0006$	$t_6(2) = 7.21$, $P = 0.043$
Human stimuli	$t_1(11) = 19.42$, $P = 4 \times 10^{-10}$	$t_2(11) = 4.36$, $P = 0.0006$	$t_3(11) = 2.06$, $P = 0.031$	$t_4(11) = 4.32$, $P = 0.0006$	$t_5(11) = 1.05$, $P = 0.16$	$t_6(11) = 5.51$, $P = 9 \times 10^{-5}$