Title
Humans Discriminate Individual Large-Billed Crows and Individual Cats by Their Respective Vocalizations

Journal
International Journal of Comparative Psychology, 34(0)

ISSN
0889-3675

Authors
Schalz, Sabrina
Meekings, Sophie
Dickins, Thomas E.

Publication Date
2021

INTERNATIONAL JOURNAL OF COMPARATIVE PSYCHOLOGY

2021, 34
Heather M. Hill
Editor
Peer-reviewed

Humans Discriminate Individual Large-Billed Crows and Individual Cats by Their Respective Vocalizations

Sabrina Schalz[1], Sophie Meekings[2], and Thomas E. Dickins[1]

[1] Middlesex University, Department of Psychology, London, UK
[2] Newcastle University, Department of English Literature, Language and Linguistics, Newcastle, UK

Previous research has shown that human adults can easily discriminate 2 individual zebra finches (*Taeniopygia guttata*) by their signature songs, struggle to discriminate 2 individual rhesus monkeys (*Macaca mulatta*) by their calls, and are unable to discriminate 2 individual dogs (*Canis familiaris*) by their barks. The purpose of the present experiment was to examine whether acoustic discrimination of individual nonprimate heterospecifics is limited to species producing stereotyped signature songs or whether it is possible with the vocalizations of other species as well. This was tested here with the calls of individual large-billed crows (*Corvus macrorhynchos*) and the meows of individual domestic cats (*Felis catus*) using a forced-choice same-different paradigm. Results show a high discrimination accuracy without prior training, although the scores obtained here for both species were lower than those in the zebra finch discrimination task. Discrimination accuracy of cat voices decreased when mean pitch was equalized between individuals but was still possible without this cue. The removal of formant frequencies did not influence the discrimination, and there was no significant performance improvement across trials. These experiments suggest that individual acoustic discrimination is possible not only with species producing signature songs but also with unlearned vocalizations of both birds and nonhuman mammals.

Keywords: cat, large-billed crow, voice discrimination, voice perception

Many animals can discriminate individual voices not only from members of their own species but from those of other species as well. Individual discrimination and recognition of human voices has been shown for several nonhuman animals: (1) Carrion crows (*Corvus corone*) and cheetah (*Acinonyx jubatus*) can discriminate between familiar and unfamiliar human voices (Leroux et al., 2018; Wascher et al., 2012), (2) domestic cats (*Felis catus*) and dogs (*Canis familiaris*) can tell the difference between the voice of their owner and that of an unfamiliar human (Adachi et al., 2007; Saito & Shinozuka, 2013), and (3) rhesus monkeys (*Macaca mulatta*) can match familiar human voices to the correct faces (Sliwa et al., 2011).

This is perhaps not surprising given the structural and functional similarities in vertebrate sound production. The source signal (the fundamental frequency F0, perceived as pitch) is produced when the vocal folds of the larynx vibrate. Its frequency depends on the size of the vocal folds and how fast the glottis opens and closes. The source signal is then filtered in the vocal tract. The vocal tract has its own resonant properties determined by its shape, and some frequencies are amplified while others are decreased (resulting in formant frequencies). The result is a vocalization that contains pitch and formant frequencies influenced by the individual's larynx size and vocal tract shape (Taylor & Reby, 2010; Titze, 1994) and that consequently contributes to vocal identification. This sound production mechanism is similar to avian vocalization, which also relies on the interplay of respiration, the vocal organ (the syrinx instead of the larynx), the vocal tract, and oscillating tissue to produce sounds (Riede & Goller, 2010b).

Despite this, human adults were previously thought to struggle with the acoustic discrimination of individual heterospecifics, as, during infancy, our auditory perception is tuned towards speech over nonhuman acoustic vocalizations. Object categorization is supported by both lemur vocalizations and human speech in 3- and 4-month-old infants but only by human speech in 6-month-olds (Ferry et al., 2013). Neonates equally prefer human speech and the vocalizations of rhesus monkeys over synthetic sounds but prefer human speech over rhesus vocalizations and synthetic sounds three months later (Vouloumanos et al., 2010). While 6- and 12-month-old human infants can accurately discriminate individual rhesus monkeys by their voices, the

Please send correspondence to Sabrina Schalz. Email: ss3903@live.mdx.ac.uk

discrimination accuracy was already lower among the older infants, and adults performed only slightly above chance level (Friendly et al., 2014). Discrimination of individual dogs by their barks had been unsuccessful with human adults (Molnár et al., 2006), even though other dogs can perceive individuality in barks (Molnár et al., 2009). Contrary to these findings, a more recent experiment with zebra finch songs showed high discrimination accuracy for adult participants (Schalz & Dickins, 2021). Zebra finches produce learned signature songs (Miller, 1979) with distinct patterns of pitch frequencies (pitch contour). Discrimination success dropped when pitch contour was removed but was still well above chancel level (Schalz & Dickins, 2021). It was unclear whether discrimination success was due to similarities between learned birdsong and human speech or whether it could be generalized to the vocalizations of other species as well.

To further investigate individual discrimination of heterospecifics by human adults, we first replicated the experiment in a small study with calls from another bird, the large-billed crow (*Corvus macrorhynchos*), to confirm that individual discrimination is possible with avian vocalizations other than signature songs. We then conducted a larger experiment with domestic cat vocalizations to examine whether individual discrimination was in fact possible with nonhuman mammalian voices and then examined which acoustic cues were relevant for the task.

Humans primarily rely on mean pitch to discriminate others by their voices, followed by formant frequency and formant dispersion (Baumann & Belin, 2010). Although sensitivity to F0 in phoneme categorization tasks varies between individuals (Kong & Edwards, 2016), these cues are likely important for heterospecific voice perception as well. The isolation calls produced by domestic kittens differ intra-individually in pitch, formant frequency, and formant dispersion (Scheumann et al., 2012). The signature voice system by which large-billed crows recognize flock members also relies on various pitch parameters, including mean frequency (Kondo et al., 2010). Both species should therefore produce vocalizations with the necessary acoustic identity cues.

To our knowledge, individual discrimination of cat voices has not yet been tested with human participants, but acoustic perception of emotional states has been addressed in multiple studies. One experiment found that human discrimination of emotional state conveyed through meows (waiting for food, isolation, and brushing against a human) is inaccurate (Prato-Previde et al., 2020). Another found that human listeners successfully discriminate between meows produced in two different contexts, although cat owners performed better than non-cat owners. The authors hypothesized that discrimination was enabled by pitch contour differences (Schötz & van de Weijer, 2014). Human discrimination between meows produced in five different contexts was only slightly above chance level, though previous experience with cats again had a positive effect on discrimination accuracy (Nicastro & Owren, 2003).

Individuals Discrimination of Large-Billed Crows

Materials and Method

Subjects

Participants (*N*=50; 32 female, 17 male) were students and staff at Middlesex University between the ages 18 to 50 who did not report hearing problems and gave informed consent. First-year psychology students received credit points for participation (*N*=29).

Stimuli

Stimuli were nine calls recorded from each of two large-billed crows (both female, 4 years old). Calls were recorded in the same context and environment by S. S. at Keio University, Tokyo. All acoustic measurements were made in Praat version 6.0.49 (Boersma & Weenink, 2020; see Table 1 for acoustic parameters and Appendix Figures 5 and 6 for sample spectrograms).

Table 1

Acoustic Parameters of the Nine Selected Calls of Each Crow

	Crow A mean	Crow A *SD*	Crow B mean	Crow B *SD*
Duration per call (ms)	263.5	0.005	265.1	0.001
Intensity per call (dB)	77.6	0.4	77.6	0.3
Pitch per call (Hz)	482.4	88.9	355.3	46.2
Frequency of the first formant (Hz)	1388.4	38.7	1462.2	26.1
Frequency of the second formant (Hz)	2408.5	165.3	3053.3	137.7

Note. Frequency range was set at 50Hz to 1,000Hz for pitch. Data is reported out to one decimal point based on measurements reported by Praat (Boersma & Weenink, 2020). Measurements only include one decimal place instead of two to most appropriately reflect the level of accuracy given by measurements of natural vocalizations and their gradual onset and offset. This did not affect the experiment.

Apparatus

Data collection took place in person in February 2020. The experimental task and participant background questionnaire were presented in PsychoPy version 3.2 (Peirce et al., 2019) on a desktop computer in a quiet room. Participants' responses were collected by PsychoPy. Stimuli were played through over-ear headphones.

Procedure

Discrimination accuracy was assessed with a forced-choice same-different paradigm (Pisoni & Lazarus, 1974) with a total of 40 trials (as in Friendly et al., 2014). Each trial consisted of two calls, either produced by the same individual (*same-trial*) or by two different individuals (*different-trial*). Pairs and trials were randomly assigned by PsychoPy, meaning two calls were randomly combined each trial. This means whether a trial contained two calls from the same individual (same-trial) or one call per individual (different-trial) was random and not balanced across the experimental session for each participant. While counter-intuitive at first, this was done on purpose to avoid predictability, keeping participants from either recognizing specific pairs or using the procedure of exclusion (i.e., "I've mostly heard different-pairs so far, so the rest must be mostly same-pairs"). The number of same- and different-trials was however balanced across participants (1,003 trials were same-trials and 997 were different-trials). After listening to both calls, participants were asked whether they thought the call was produced by the same individual, to which they could respond "yes" and "no." Participants did not receive training before or feedback during the experiment.

Participants' responses were divided into four response types: *hit* (yes on a same-trial), *miss* (no on a same-trial), *correct reject* (no on a different-trial), and *false alarm* (yes on a different-trial). The proportions of hit responses on same-trials and false alarm responses on "different"-trials were used to calculate the discrimination sensitivity index *d'* (Stanislaw & Todorov, 1999) in R version 3.6.1 (R Core Team, 2020) using the R package psyphy (Knoblauch, 2014). The resulting *d'* scores indicate the discrimination accuracy on a continuous scale from 0 (chance level, no discrimination) to 5.94 (perfect discrimination). The *d'* scores are an appropriate measurement for this experiment because they are less susceptible to participants' response biases than other methods, such as the percentage of correct responses per participant (Stanislaw & Todorov, 1999). The *d'* scores cannot be calculated with absolute values for hit rates and false alarm rates, so three hit rates and five false-alarm rates were corrected following Equation 1 as described in (Snodgrass & Corwin, 1988).

$$1) \quad corrected\ rates = \frac{0.5 + responses\ (either\ hit\ or\ false\ alarm)}{1 + number\ of\ trials\ (either\ same\ or\ different)}$$

Statistical Analyses

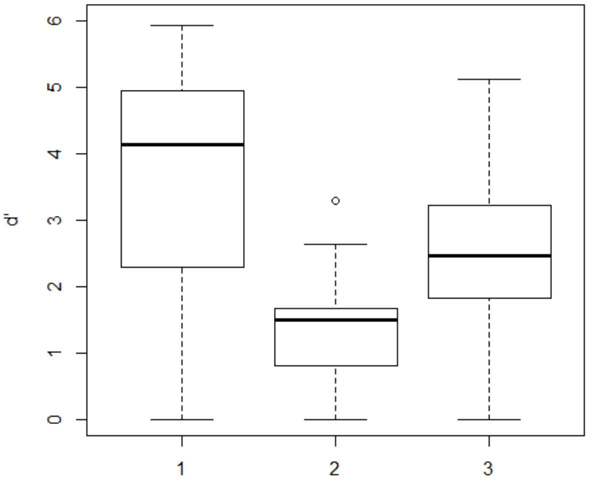
Whether *d'* scores were significantly above chance level was tested with a one-sample *t*-test (*mu* = 0). The *d'* scores were also compared with those obtained in the previous experiment with zebra finch songs (Schalz & Dickins, 2021) using a Mann-Whitney *U* test. Whether discrimination accuracy improved over the course of the trials was examined with a Mann-Kendall trend test with the R package Kendall (McLeod, 2011). Data for this analysis consisted of the percentage of correct responses (either hit or correct reject) for a given trial out of all responses for that trial (one response per participant per trial).

Results

The average *d'* score was 2.48 (*SD* = 1.09, 95% *CI* [2.17, 2.78]), and *d'* scores were significantly above chance level *t*(49) = 15.86, *p* < .01). Compared with the *d'* scores reported for the discrimination of zebra finches (Schalz & Dickins, 2021), *d'* scores obtained in the present experiment were significantly below those obtained with natural zebra finch song (*W* = 669, *d* = 0.9, *p* < .01), and significantly above those obtained for song without pitch contour (*W* = 1024.5, *p* < .01, *d* = 1.22,; see Figure 1).

Figure 1

Comparision of d' Scores Between Conditions

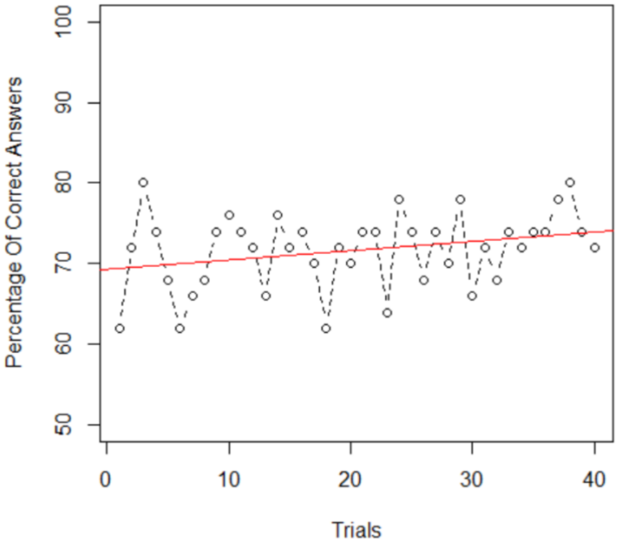


Note. Comparision of *d'* scores between (1) natural zebra finch songs, (2) zebra finch songs without pitch contour (Schalz & Dickins, 2021), and (3) crow calls (3).

The Mann Kendall Trend test showed a non-significant, although positive trend across all trials (τ = 0.19; see figure 2).

Figure 2

Percentage of Respondents who Gave a Correct Answer (either Hit or Correct Reject) in Each Trial



Note. Percentage was computed as the sum of correct answers for each trial, with a linear regression line indicating overall trend.

Individual Discrimination of Domestic Cats

The experiment consisted of 3 conditions, the first using natural stimuli to test the discrimination of natural meows, followed by two smaller conditions using modified stimuli to test the relevance of pitch and formants as identity cues. Differences between conditions are detailed in on page 2 of this paper "Materials and Method – Stimuli." All conditions have been approved by the Middlesex University Psychology Research Ethics Committee.

Materials and Method

Subjects

A total of 100 adults participated, 50 in Condition 1 and 25 in each of Conditions 2 and 3 (see Table 2 for details). Volunteers were recruited on Twitter, Facebook, and among Middlesex students. One first-year psychology student received credit points for participation. Participants had not contributed recordings of their own cat and did not consider themselves to have a hearing impairment. Participants over the age of 65 were excluded from the experiment as hearing naturally declines at this age (Harris et al., 2008). All participants gave informed consent. Data were collected from January 2 to April 21, 2021.

Table 2

Number of Participants (N) of Each Sex, Mean Age in Years, and Share of Participants with Cat Experience per Condition

Condition	*N*	Sex	Mean Age (*SD*)	Cat Experience (%)
1	50	37 female, 10 male, 3 other	31.5 (8.1)	42 (84%)
2	25	21 female, 3 male, 1 other	35.8 (10.4)	22 (88%)
3	25	20 female, 4 male, 1 other	30 (7.9)	18 (72%)

Note. Cat Experience was defined as "having lived with a cat for at least a year."

Stimuli

In Condition 1, stimuli consisted of natural meows of 6 domestic cats (female, 1-9 years old). All recordings had been submitted by volunteers who had been asked to record 20 meows of their cats in any context. Those recordings with the highest audio quality were selected for the experiment. Acoustic parameters were analyzed in Praat (see Table 3 and Appendix, Figures 7 - 12; Boersma & Weenink, 2020). Meows were selected in the order they appeared until a total of 20 meows with sufficiently high audio quality had been reached or until there were no more meows in the recording. Cats were matched into pairs based on similarity of "meow style" to avoid discrimination based on difference in emotional state of the cat. Fourteen meows per cat were included to match the maximum meows of the cat with the fewest meows, resulting in a total of 84 meows. Amplitude was equalized for all stimuli using Audacity version 2.4.2 (Audacity Team, 2021). Length was not equalized across all stimuli as meow length naturally differs inter- and intraindividually, but pairs were matched to be as similar in length as possible, and the longer meow was shortened to match its counterpart in the specific pair. Duration could therefore not be used as a discrimination cue.

Table 3

Pairs, Numbers of Meows, and Acoustic Parameters of Meows for Each Cat as Measured in Praat (Boersma & Weenink, 2020)

Pairs	**Pair 1**		**Pair 2**		**Pair 3**	
Cat	**A**	**B**	**C**	**D**	**E**	**F**
Mean duration in seconds (*SD*)	0.7 (0.05)	0.7 (0.05)	1 (0.3)	1 (0.3)	0.4 (0.2)	0.4 (0.2)
Mean intensity in dB (*SD*)	60.1 (0.4)	60.4 (0.4)	60.5 (0.5)	60.7 (0.5)	60.5 (0.6)	60.2 (0.6)
Mean pitch in Hz (*SD*)	716.4 (49.7)	650.8 (32.9)	587.3 (66.1)	530.3 (26)	651.6 (80.3)	815.4 (102.7)
Mean F1 in Hz (*SD*)	1749.4 (153.7)	1463.4 (217.8)	1395.8 (177.6)	1751.9 (170.3)	1659.7 (161)	1616.2 (251.1)
Mean F2 in Hz (*SD*)	2697.4 (111)	2725.4 (204.5)	2643.1 (248)	3018 (159.8)	2941 (193.1)	3039.2 (288.8)
Mean Formant Dispersion F2-F1 in Hz (*SD*)	948 (136.3)	1262.1 (170.8)	1247.4 (135.6)	1266.1 (161.9)	1281.4 (149.8)	1423 (286.8)

Note. Settings: spectrogram window length 0.01, pitch 75Hz - 3kHz, 4 formants, max formant 8kHz, formant window length 0.01. Measurements only include one decimal place instead of two to most appropriately reflect the level of accuracy given by measurements of natural vocalizations and their gradual onset and offset. This did not affect the experiment.

For Conditions 2 and 3, meows were manipulated in Praat (Boersma & Weenink, 2020). In Condition 2, pitch was equalized across all stimuli by removing existing pitch points and replacing them with a single new point at 600Hz in Praat. Pitch contour was already flat in the natural stimuli, so this did not significantly change during the manipulation. In addition to the pitch equalization, formant frequencies were removed in Condition 3 by adding a low-pass filter at 1,000Hz using the Praat plugin "Vocal Toolkit" (Corretge, 2020), following the approach previously used by Xu et al. (2013). This is well above the equalized pitch frequency of 600Hz and below the natural formant frequencies (see Table 3).

Apparatus

Due to lockdown restrictions prohibiting in-person data collection, the experiment was designed and presented online using the software PsychoPy version 3.2 (Peirce et al., 2019) and hosted on the server pavlovia.com. Participants were asked to complete the experiment in a quiet room while wearing headphones.

Procedure

Discrimination was tested with the same forced-choice same-different paradigm described previously in the large-billed crow experiment, in a total of 42 trials. Each participant received 21 same-trials and 21 different-trials. Each session included seven meows per cat, and all participants heard all six cats throughout their session. The order of pairs was randomized in each trial. Before the experiment, participants were asked about their age, sex, and whether they had ever lived or worked with a cat for more than a year in total.

Analysis

The analysis was conducted in R version 4.0.2 (R Core Team, 2020). The data in Conditions 1 and 2 were normally distributed (as tested with a Shapiro test), and a one-sample *t*-test was used to examine whether *d'* scores were significantly above chance level (*mu* = 0). For Condition 3, a Mann-Whitney *U* test was used instead. Whether *d'* scores differed significantly between conditions was tested with a Kruskal-Wallis test and a Wilcoxon signed-rank test as posthoc. Whether *d'* scores in Condition 1 differed from those obtained with natural zebra finch songs and large-billed crow calls was tested with a Wilcoxon signed-rank test.

Whether discrimination improved over time (percentage of correct answers per trial pooled from all participants) was assessed with a linear regression model lm(percentage correct ~ trial number) for all conditions.

Results

In all three conditions, *d'* scores were significantly above chance level (Condition 1: *t*(49) = 25.24, *p* < .01; Condition 2: *t*(19) = 10.00, *p* < .01; Condition 3: *V* = 231, *p* < .01; see Table 4).

Table 4

Outcomes for Each Condition

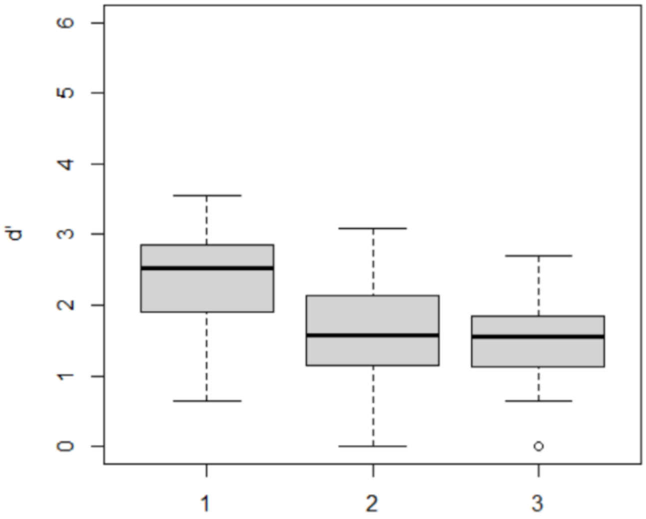
Condition	Mean *d'*	*SD d'*	95% *CI*	Min. *d'*	Max. *d'*	% Correct
1	2.38	0.66	2.19 - 2.56	0.65	3.56	70.3
2	1.59	0.71	1.31 - 1.87	0.00	3.08	61.6
3	1.36	0.73	1.05 - 1.66	0.00	2.69	59.2

Note. Displays mean *d'* scores, SD for *d'* Scores, 95% confidence interval, highest and lowest *d'* scores, and the overall percentage of correct trials for each condition. the highest possible score was 5.94.

In Condition 1, *d'* scores were significantly higher than those in Conditions 2 and 3, respectively (*p* < 0.01), but *d'* scores did not differ significantly between Conditions 2 and 3 (see Figure 3). They were also significantly lower than those from the experiment with natural zebra finch songs (*W* = 1894.5, *p* < 0.01; Schalz & Dickins, 2021) but did not differ significantly from those obtained with large-billed crow calls (*W* = 1299.5, *n. s.*).

Figure 3

Comparision of d' Scores Between Conditions



Note. Comparision of *d'* scores between (1) natural cat meows, (2) cat meows with equalized pitch, and (3) cat meows with equalized pitch and no formant frequencies.

Table 5

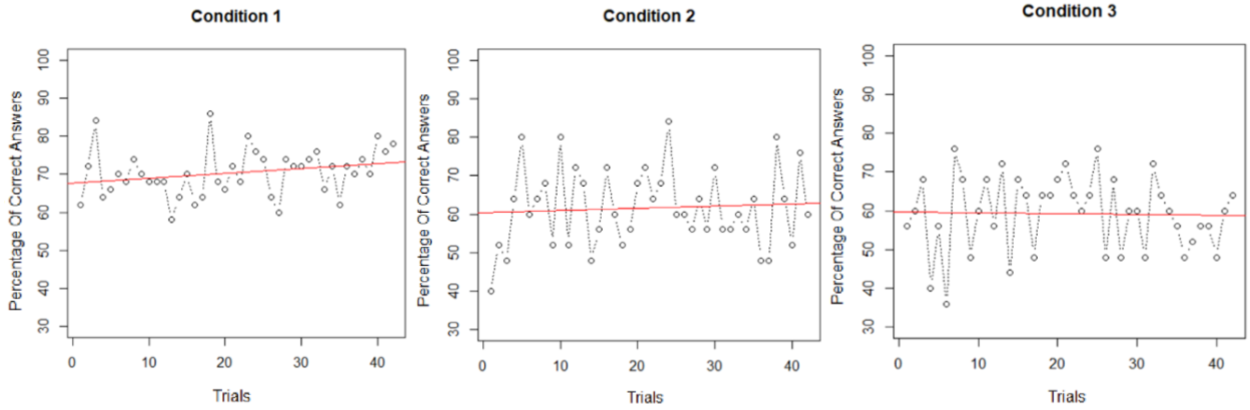
Median Hit and False Alarm Rates for Each Condition

	Median hit rate	***SD* hit rate**	**Median false alarm rate**	***SD* false alarm rate**
Condition 1	0.80	0.11	0.35	0.12
Condition 2	0.57	0.14	0.38	0.18
Condition 3	0.66	0.12	0.42	0.15

There was no significant trend across trials in any of the three conditions (see Figure 4).

Figure 4

Percentage of Respondents Who Gave a Correct Answer (either Hit or Correct Reject) in Each



Note. The sum of correct answers for each trial. The linear regression line indicates overall trend.

Discussion

These results show that humans are able to discriminate not only individual zebra finches by their signature songs but also large-billed crows by their ka-calls and domestic cats by their meows. However, discrimination accuracy was significantly higher for zebra finch song (*M d'* = 3.68; Schalz & Dickins, 2021) than for large-billed crows (*M d'* = 2.48) or cats (*M d'* = 2.38). This may be due to the more stereotyped pattern in natural zebra finch songs, which likely facilitates the discrimination compared to that of unlearned vocalizations like crow calls and cat meows.

As with the zebra finch songs, discrimination accuracy of cat meows decreased significantly when pitch was equalized between individuals (*M d'* = 1.59). This shows that pitch is an important cue for individual discrimination of both species. While the false alarm rate stayed at the same level after pitch was equalized, the hit rate declined considerably and was close to chance level (see Table 5). This suggests that pitch is important to tell individuals together (i.e., recognize multiple meows belong to the same individual) but not to tell individuals apart. As discrimination was still possible without it, pitch is not the only relevant cue. While formant frequencies are important cues for human voice discrimination (Baumann & Belin, 2010), accuracy did not significantly decline after formant frequencies were removed from the meows (*M d'* = 1.36), so this does not seem to be the case for individual cat discrimination, and another cue seems to be used instead. This other cue could be timbre, which is what allows listeners to distinguish sounds that are otherwise equal in their acoustic properties, including in their pitch and formant frequencies (Town & Bizley, 2013). For instance, human listeners use a combination of pitch and timbre for voice gender categorization and can perform this task when only timbre is available (Pernet & Belin, 2012). Another possible discrimination cue is temporal variation within each vocalization. While mean amplitude for instance was equalized for each stimulus as a whole, amplitude patterns within a single stimulus were not. Whether such within-stimulus patterns support individual discrimination could be tested with stimuli consisting only of repetitions of a single cycle from the individual's vocalization (A. Jorschick, personal communication, March 3, 2021).

There was no significant learning trend across trials, and there seems to be no meaningful improvement of discrimination accuracy throughout the experiment. The same was observed for the discrimination of individual zebra finches (Schalz & Dickins, 2021). It is however conceivable that accuracy would eventually improve after additional practice. It is also possible that some participants' individual accuracy improved over the course of these trials, which would not necessarily appear in the pooled performance analysed here.

Previous studies had suggested adult humans could not discriminate individual nonhuman voices well. With rhesus monkey vocalizations, adults had a mean *d'* score of 0.37 and 6-month-old infants had a mean score of 0.79 (Friendly et al., 2014), while the discrimination of individual dogs had been entirely unsuccessful with human adults (Molnár et al., 2006). However, dogs are able to discriminate each other by their barks (Molnár et al., 2009), and rhesus monkeys can match the voices of familiar conspecifics and humans to their faces (Sliwa et al., 2011), suggesting that the low discrimination success in humans is not due to a lack of voice cues in these species' vocalizations. These results may instead be due to methodological differences. For instance, the R package psycho will produce different *d'* scores for the same hit and false alarm rates than the package psyphy used here. As such, *d'* scores should be seen relative to scores within the same experiment and should only be cautiously compared to other experiments. Friendly et al. (2014) did not specify which method they used. In addition to that, all mean *d'* score comparisons here are post hoc comparisons between participants, and, considering the high individual variation in *d'* scores, it is possible that some groups randomly performed better than others. Friendly et al. (2014) had also matched the stimuli pairs not only for mean duration but also for mean, minimum, and maximum pitch. Considering the relevance of both mean pitch and pitch contour, this may have added an additional difficulty to the discrimination, and the discrimination may have been easier with other rhesus monkey voice pairs. Molnár et al. (2006) chose dog barks as voice stimuli because previous studies had shown that humans struggle to perceive acoustic cues in chaotically noisy sounds. It is possible that humans would be able to discriminate individual dogs by other, less chaotic vocalizations, such as howls or whines.

Based on the previous findings with zebra finches (Schalz & Dickins, 2021) and the findings reported here, we should expect individual discrimination to be possible for humans with any nonhuman species that produces vocalizations within human hearing ranges and that differ inter-individually in their fundamental frequencies. As the fundamental frequency of an individual depends on the size of the vocal folds (Titze, 1994), this should generally apply to species whose vocal sound production involves the vocal folds or a comparable vibrating soft tissue that can vary in size interindividually: mammals (Taylor & Reby, 2010; Titze, 1994), birds (Riede & Goller, 2010a), and at least some reptiles (Moore et al., 1991; Young et al., 1995). Put another way, a voice is still a voice when it belongs to a nonhuman individual, and we should expect human listeners to perceive them just like they would a human voice. While this sounds anything but surprising, it contradicts the previous findings on individual heterospecific discrimination by humans (as discussed above). It also questions the assumption that the human vocal communication system is unique: While voices are certainly less complex than language, they are part of vocal communication systems, and bidirectional voice perception (see Introduction for examples of individual human discrimination by nonhuman animals) constitutes one less divide between human and nonhuman vocalizations. The findings reported here therefore contribute to the effort of moving away from assumed human uniqueness and towards a less biased approach of "treating humans as an equal species in vocal-communication research" (Prat, 2019, p.1).

Acknowledgments

We thank Ei-Ichi Izawa for enabling the incidental recording of the crow calls during an unrelated project (Schalz & Izawa, 2020) that was carried out as part of the "Thesis@Keio" programme at Keio University and financially supported by JSPS KAKENHI #17H02653 and Keio University Grant-in-Aid for Innovative Collaborative Research Projects #MKJ1905 to E.-I. I. We also thank Annett Jorschick for a helpful discussion during the planning of condition 3 of the cat experiment. We are grateful to all volunteers who have provided recordings of their cats, as well as those who have participated in the experiments.

Data Availability

R script, data, and stimuli used in the crow experiment are freely available here:
https://doi.org/10.6084/m9.figshare.c.5079260.v1

R script, data, and stimuli used in the cat experiment are freely available here:
https://doi.org/10.6084/m9.figshare.c.5529099.v1

Author Contributions

S. S. designed and performed the experiments, conducted the analysis, and led the writing of the manuscript. S. M. added to the design of Experiment 2 Condition 3 and contributed to the draft. T. D. supervised the project, added to the argument, and contributed to the draft.

References

Adachi, I., Kuwahata, H., & Fujita, K. (2007). Dogs recall their owner's face upon hearing the owner's voice. *Animal Cognition*, *10*(1), 17–21. https://doi.org/10.1007/s10071-006-0025-8

Audacity Team. (2021). *Audacity®* (2.4.2). https://www.audacityteam.org/about/citations-screenshots-and-permissions/

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research*, *74*(1), 110–120. https://doi.org/10.1007/s00426-008-0185-z

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (6.1.16). https://www.fon.hum.uva.nl/praat/

Corretge, R. (2020). *Praat Vocal Toolkit*. http://www.praatvocaltoolkit.com

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15231–15235. https://doi.org/10.1073/pnas.1221166110

Friendly, R. H., Rendall, D., & Trainor, L. J. (2014). Learning to differentiate individuals by their voices: Infants' individuation of native- and foreign-species voices. *Developmental Psychobiology*, *56*(2), 228–237. https://doi.org/10.1002/dev.21164

Harris, K. C., Mills, J. H., He, N. J., & Dubno, J. R. (2008). Age-related differences in sensitivity to small changes in frequency assessed with cortical evoked potentials. *Hearing Research*, *243*(1–2), 47–56. https://doi.org/10.1016/j.heares.2008.05.005

Knoblauch, K. (2014). *psyphy: Functions for analyzing psychophysical data in R*. https://cran.r-project.org/web/packages/psyphy/index.html

Kondo, N., Izawa, E. I., & Watanabe, S. (2010). Perceptual mechanism for vocal individual recognition in jungle crows (*Corvus macrorhynchos*): Contact call signature and discrimination. *Behaviour*, *147*(8), 1051–1072. https://doi.org/10.1163/000579510X505427

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

Leroux, M., Hetem, R. S., Hausberger, M., & Lemasson, A. (2018). Cheetahs discriminate familiar and unfamiliar human voices. *Scientific Reports*, *8*(1), 1–6. https://doi.org/10.1038/s41598-018-33971-1

McLeod, A. I. (2011). *Kendall rank correlation and Mann-Kendall trend test* (2.2 https://cran.r-project.org/web/packages/Kendall/index.html

Miller, D. B. (1979). The acoustic basis of mate recognition by female Zebra finches (*Taeniopygia guttata*). *Animal Behaviour*, *27*(2), 376–380. https://doi.org/10.1016/0003-3472(79)90172-6

Molnár, C., Pongrácz, P., Dóka, A., & Miklósi, Á. (2006). Can humans discriminate between dogs on the base of the acoustic parameters of barks? *Behavioural Processes*, *73*(1), 76–83. https://doi.org/10.1016/j.beproc.2006.03.014

Molnár, C., Pongrácz, P., Faragó, T., Dóka, A., & Miklósi, Á. (2009). Dogs discriminate between barks: The effect of context and identity of the caller. *Behavioural Processes*, *82*(2), 198–201. https://doi.org/10.1016/j.beproc.2009.06.011

Moore, B. A., Russell, A. P., & Bauer, A. M. (1991). Structure of the larynx of the tokay gecko (*Gekko gecko*), with particular reference to the vocal cords and glottal lips. *Journal of Morphology*, *210*(3), 227–238. https://doi.org/10.1002/jmor.1052100303

Nicastro, N., & Owren, M. J. (2003). Classification of domestic cat (*Felis catus*) vocalizations by naive and experienced human listeners. *Journal of Comparative Psychology*, *117*(1), 44–52. https://doi.org/10.1037/0735-7036.117.1.44

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology*, *3*(23). https://doi.org/10.3389/fpsyg.2012.00023

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, *55*(2), 328–333. https://doi.org/10.1121/1.1914506

Prat, Y. (2019). Animals have no language, and humans are animals too. *Perspectives on Psychological Science*, *14*(5), 885–893. https://doi.org/10.1177/1745691619858402

Prato-Previde, E., Cannas, S., Palestrini, C., Ingraffia, S., Battini, M., Ludovico, L. A., Ntalampiras, S., Presti, G., & Mattiello, S. (2020). What's in a meow? A study on human classification and interpretation of domestic cat vocalizations. *Animals*, *10*(12), 1–18. https://doi.org/10.3390/ani10122390

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org

Riede, T., & Goller, F. (2010a). Functional morphology of the sound-generating labia in the syrinx of two songbird species. *Journal of Anatomy*, *216*(1), 23–36. https://doi.org/10.1111/j.1469-7580.2009.01161.x

Riede, T., & Goller, F. (2010b). Peripheral mechanisms for vocal production in birds - differences and similarities to human speech and singing. *Brain and Language*, *115*(1), 69–80. https://doi.org/10.1016/j.bandl.2009.11.003

Saito, A., & Shinozuka, K. (2013). Vocal recognition of owners by domestic cats (*Felis catus*). *Animal Cognition*, *16*(4), 685–690. https://doi.org/10.1007/s10071-013-0620-4

Schalz, S., & Dickins, T. E. (2021). Humans discriminate individual zebra finches by their song. *Biolinguistics*, *14*(SI), 130–144. https://biolinguistics.eu/index.php/biolinguistics/article/view/718

Scheumann, M., Roser, A. E., Konerding, W., Bleich, E., Hedrich, H. J., & Zimmermann, E. (2012). Vocal correlates of sender-identity and arousal in the isolation calls of domestic kitten (*Felis silvestris catus*). *Frontiers in Zoology*, *9*(1), 36. https://doi.org/10.1186/1742-9994-9-36

Schötz, S., & van de Weijer, J. (2014). A study of human perception of intonation in domestic cat meows. *Proceedings of the International Conference on Speech Prosody*, 874–878. https://doi.org/10.21437/speechprosody.2014-163

Sliwa, J., Duhamel, J. R., Pascalis, O., & Wirth, S. (2011). Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1735–1740. https://doi.org/10.1073/pnas.1008169108

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Taylor, A. M., & Reby, D. (2010). The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology*, *280*(3), 221–236. https://doi.org/10.1111/j.1469-7998.2009.00661.x

Titze, I. R. (1994). *Principles of Voice Production*. Prentice Hall.

Town, S. M., & Bizley, J. K. (2013). Neural and behavioral investigations into timbre perception. *Frontiers in Systems Neuroscience*, *7*(88), 1–14. https://doi.org/10.3389/fnsys.2013.00088

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. https://doi.org/10.1111/j.1467-8624.2009.01412.x

Wascher, C. A. F., Szipl, G., Boeckle, M., & Wilkinson, A. (2012). You sound familiar: Carrion crows can differentiate between the calls of known and unknown heterospecifics. *Animal Cognition*, *15*(5), 1015–1019. https://doi.org/10.1007/s10071-012-0508-8

Xu, M., Homae, F., Hashimoto, R., & Hagiwara, H. (2013). Acoustic cues for the recognition of self-voice and other voice. *Frontiers in Psychology*, *4*(735), 1–7. https://doi.org/10.3389/fpsyg.2013.00735

Young, B. A., Sheft, S., & Yost, W. (1995). Sound production in *Pituophis melanoleucus* (Serpentes: Colubridae) with the first description of a vocal cord in snakes. *Journal of Experimental Zoology*, *273*(6), 472–481. https://doi.org/10.1002/jez.1402730604

Financial conflict of interest: No stated conflicts.
Conflict of interest: No stated conflicts.

Submitted: *June 7th, 2021*
Resubmitted: *July 26th, 2021*
Accepted: *July 31st, 2021*