ABC 2023, 10(2):144-162
Animal Behavior and Cognition
DOI: https://doi.org/10.26451/abc.10.02.04.2023

.

Wild Carrion Crows (*Corvus corone*) Autonomously Respond to Speech but Show No Difference in Their Response to a Local and a Foreign Language

Sabrina Schalz

Middlesex University, Department of Psychology, London, UK

Email: sabrina.schalz@outlook.com

Citation – Schalz, S. (2023). Wild carrion crows (*Corvus corone*) autonomously respond to speech but show no difference in their response to a local and a foreign language. *Animal Behavior and Cognition, 10*(2), 144-162. https://doi.org/10.26451/abc.10.02.04.2023

Abstract – Eavesdropping on the vocalizations of other species can be beneficial for wildlife to avoid predator encounters, including encounters with humans. Wild-caught large-billed crows in Tokyo responded more to playback of a foreign language than to Japanese without any training or rewards provided in the experiment, suggesting habituation to the local language. Here I tested the response of wild carrion crows in the UK to playback of a foreign language (Vietnamese), the local language (English), and non-speech control vocalizations (pigeon and parakeet vocalizations) to examine whether wild crows eavesdrop on speech. Playback experiments were conducted in two cities that differ in their population size and linguistic diversity (London and Milton Keynes, Buckinghamshire), to understand the role of exposure frequency to humans and to different languages in their response to speech. The crows autonomously responded with increased flight behaviors to human speech compared to non-speech control vocalizations. However, unlike previously shown for large-billed crows, the carrion crows did not respond differently to the two languages. It remains to be understood whether eavesdropping on speech provides any benefit to the animals, particularly urban individuals with frequent exposure to humans.

Keywords – Speech perception, language discrimination, playback, flight response

---

Eavesdropping on heterospecific vocalizations may alert individuals to the presence of a predator (Magrath et al., 2015), and several species have been shown to respond to human vocalizations as well. For example, urban herring gulls (*Larus argentatus*) show the same decrease in body temperature (physiological stress response) to human shouting as they do to conspecific alarm calls (Di Giovanni et al., 2022). Captive carrion crows (*Corvus corone*) and wild Western Australian magpies (*Gymnorhina tibicen dorsalis*) can discriminate between familiar and unfamiliar human voices and respond more to the latter, possibly to assess whether a given person poses a threat (Dutour et al., 2021; Wascher et al., 2012). Speech playback in the Santa Cruz Mountains (usually not accessible to human visitors) led to an avoidance of the area by mountain lions (*Puma concolor*), as well as reduced activity in bobcats (*Lynx rufus*), striped skunks (*Mephitis mephitis*) and Virginia opossums (*Didelphis virginiana*) (Suraci et al., 2019).

One feature that distinguishes different languages and may enable language discrimination is prosody, which refers to speech rhythm based on parameters such as word-stress, tone, and intonation (Hirst & Di Cristo, 1998). For instance, romance languages such as Spanish are classified as syllable-timed, Germanic languages such as English are stress-timed, and Japanese is mora-timed (Abercrombie, 1967; Ladefoged & Johnson, 2014; Pike, 1945). French newborn human infants can discriminate between

languages based on prosodic differences; they can discriminate stress-timed English from mora-timed Japanese but cannot discriminate stress-timed English from stress-timed Dutch (Nazzi et al., 1998).

Building on this work, cotton-top tamarin monkeys (*Saguinus oedipus*) have been shown to discriminate Dutch and Japanese in a habituation-dishabituation experiment, but not when the sentences were played backwards. They also successfully (although to a lesser extent) completed this task with artificial stimuli in which lexical and phonetic information was removed from the original sentences, leaving only prosodic cues (Ramus et al., 2000). Discrimination of Dutch and Japanese has also been shown for rats (*Rattus norvegicus*), who were able to discriminate the forward but not backward-played natural and synthetic speech stimuli used by Ramus et al. (2000) (Toro et al., 2003). Among domestic animals, autonomous language discrimination was recently reported in dogs (*Canis familiaris*), who showed a novelty preference for the language not spoken in their home (Mallikarjun et al., 2022), similar to the novelty response of the large-billed crows when hearing Dutch. Older domestic dogs show more pronounced differences in neural activity patterns when hearing a familiar and an unfamiliar language, further highlighting the role of natural language exposure (Cuaya et al., 2021).

Among avian species, language discrimination has been shown in Java sparrows (*Lonchura oryzivora*), who learn to discriminate Chinese and English speech and then generalize this knowledge to new sentences (Watanabe et al., 2006). Large-billed crows (*Corvus macrorhynchos*) wild-caught in Tokyo showed an increased response to playback of Dutch (a foreign language) than to Japanese (the local language), without receiving any training or rewards in the experiment (Schalz & Izawa, 2020). To my knowledge, this was the first study showing an untrained and unincentivized differential response to a foreign and a local language in a wild or wild-caught animal. The relatively higher response to Dutch than to Japanese suggests that they had listened to speech prior to the experiment and were already familiar with Japanese, but not with Dutch. It remains to be investigated whether the crows engage in this behavior to reduce the potential risk of human presence, and whether they began eavesdropping on speech in the wild, or after being caught.

The aim of the present experiment was to test whether Carrion crows also respond autonomously to speech, and whether they also respond more to a foreign language than the local language. To confirm whether crows show this behavior prior to capture, the experiment was conducted with wild individuals. The experiment was conducted at three sites differing in human population density levels and linguistic diversity, to test whether the response to speech is correlated with the exposure to humans, and whether the response to speech is negatively correlated with the exposure to the local language. To keep results as comparable as possible to previous research on language discrimination in nonhuman animals, I used the same methodology as the laboratory experiments where possible and adapted it to field conditions where needed. As conditions in the field are different than in the previous laboratory-based experiments and behaviors are expected to differ from the previous experiments (e.g., flight is an option for wild crows but not for captive crows), the analysis is primarily exploratory while the rationale behind the experiment is based on the hypotheses outlined above.

Methods

Ethics Statement

The experiment was approved by the Middlesex University Ethics Committee (application #15200).

Study Sites

The experiment was conducted at three sites: Hampstead Heath (London 2021), the Loughton football field (suburban Milton Keynes 2022), and the Milton Keynes train station square (urban Milton Keynes 2022). Greater London is a highly urbanized city with a population of 8.8 million and a population density of 5,600/km^2, while Milton Keynes is considerably smaller at 287,000 inhabitants and a population

density of 930/km^2 (Office for National Statistics, 2022). In addition to their difference in population size, London is also considerably more multilingual: only 78% of London households use English as their main language, compared to 89% in Milton Keynes (Office for National Statistics, 2011). These differences in population density and linguistic diversity allow us to examine the potential role of exposure frequency to humans, as well as the role of exposure to the majority language (in this case English) compared to other languages.

Subjects

Subjects were wild carrion crows found on the three experimental sites. Individuals could not be visually identified and ringing them would have created a confounding effect, as the experience of being captured by a human might influence their response to human speech. The largest group size during any given trial was 10 individuals at the London site, four at the suburban Milton Keynes site, and six at the urban Milton Keynes site, but since it was not possible to identify individuals, it is not known what the number of different tested individuals was. The London site was 71km from the Milton Keynes sites, and the two Milton Keynes sites were 1.6km apart. While it is possible that individuals were present at two sites, it is unlikely: From 2010 to 2019, the British Trust for Ornithology ringing records show 10 out of 12 adult crows in England were recovered no further than 1km from their original ringing location 5-12 years previously. While this is a small sample size, it suggests low dispersal distances among adult carrion crows in England (Robinson et al., 2020).

Playback Stimuli

Stimuli in the London experiment consisted of 30 English sentences, 30 Vietnamese sentences, and 30 ring-necked parakeet (*Psittacula krameri*) vocalizations, each produced by three different individuals with 10 sentences/vocalizations per individual. In Milton Keynes, I used the same speech stimuli but used wood pigeon (*Columba palumbus*) recordings instead of parakeet vocalizations, as the latter are not commonly found in the wild in Milton Keynes, and crows there should not be familiar with them.

English speakers were female volunteers from southern UK regions recorded for the corpus "Crowdsourced high-quality UK and Ireland English Dialect speech data set" (SLR83) published by an unnamed author on opensrl.org (CC BY-NC-SA 4.0). Vietnamese stimuli were produced by female speakers and selected from the VIVOS Corpus produced by AILAB (Luong & Vu, 2016; CC BY-NC-SA 4.0). Stimuli sentences were chosen from the corpora to be as similar in duration as possible within and between languages (mean duration for English was 3.87s (*SD* = 0.46s), and 3.93s for Vietnamese (*SD* = 0.3s). All-female speakers were chosen because there may be differences in behavioral responses towards speakers of different sexes; wild jackdaws (*Corvus monedula*), for instance, respond with higher vigilance to male than female human voices (McIvor et al., 2022). The previous experiments on language discrimination in nonhuman animals (see Introduction) used female speakers only, so I chose female over male speakers for consistency.

Mean voice pitch of English speakers was 169Hz (*SD* = 9.16), 196Hz (*SD* = 8.27), and 232Hz (*SD* = 9.96) (pitch range setting was between 75 and 500 Hz, see supplementary materials for example spectrograms). Mean voice pitch of Vietnamese speakers was 225Hz (*SD* = 7.06), 235Hz (*SD* = 6.22), and 240Hz (*SD* = 6.93) (Boersma & Weenink, 2020). Vietnamese was chosen as the unfamiliar language because, unlike the stress-timed English, it is a syllable-timed language (Nguyễn, 1970) with some phonemes shared between the two languages and some phonemes only used in either English or Vietnamese (Tang, 2007). Additionally, Vietnamese is spoken by only 0.1% of London residents, so the likelihood of the crows having heard Vietnamese before is relatively low. East Asian languages in general are spoken by only 1.6% of residents (Office for National Statistics, 2011).

Nonhuman animal vocalizations (parakeet and pigeon vocalizations) were chosen as control stimuli over artificial sounds to compare the crows' responses to acoustic communications of vertebrates specifically, rather than a comparison between human vocalizations and artificial sounds like car horns,

which communicate a different type of ecological signal. The aim here was to look at vocalizations that naturally occur in the crows' habitats: Parakeets are common at the London study site (Hampstead Heath), and pigeons are common in Milton Keynes. They are neither predator nor prey for carrion crows, so their vocalizations should be familiar but not relevant to them.

Parakeet and pigeon recordings were taken from the xeno-canto database (www.xeno-canto.org, see details of the source files and example spectrograms in the supplementary materials). Each stimulus consisted of a single note, so that there would be no differences in pause length or number of notes per bout that could help discriminate the stimuli. The average duration for parakeet stimuli was 0.12s (*SD* = 0.01), whereas pigeon stimuli had an average duration of 0.29s (*SD* = 0.01). The parakeets had an average pitch of 2,508Hz (*SD* = 142.52), 2,274Hz (*SD* = 89.89), and 2,029Hz (*SD* = 484.67) (pitch range setting between 1 and 5 kHz), while the pigeons had an average pitch of 419Hz (*SD* = 58.74, pitch range setting between 75 and 1000Hz), 446Hz (*SD* = 35.28=, and 466Hz (*SD* = 42.1) (Boersma & Weenink, 2020).

A silent pause of 1s was added after each stimulus to separate different stimuli during playback, and all stimuli were equalized in intensity (70dB) using Audacity (Audacity Team, 2021). Stimuli were sorted into playlists with two stimuli per speaker/bird in each, a total duration of 30s for each speech playlists, and five playlists per stimuli group. The playback order in each playlist was set to random.

Apparatus and Set-Up

A Bluetooth speaker (JAM HX-P303) was hidden in shrubs or behind fence posts during each trial, to avoid the absence of any visual cues on an open field to influence the crows' behavior – if the sound is coming from a direction concealed by vegetation, they cannot visually ascertain the absence of humans. Food (Doritos™, Frito-Lay, USA) was placed as bait approximately 1m in front of the audio speaker. As I tested wild crows, it would not have been possible to gather them around the speaker without a food bait. It is possible that individuals may have shown a weaker flight response in order to not lose the provided food.

In each trial, a stimuli playlist was played from a Bluetooth-connected phone at approximately 65dB(A) at a 1m distance, matching the natural amplitude of loud speech (the normal range being 40 to 60dB, Pfitzinger & Kaernbach, 2008) while also being above the average noise levels of central London public parks (50 to 60dB, Bose & Skinner, 2009) so as to not be masked out by background noise. While I did not find natural amplitude ranges for either wood pigeons or parakeets, 65dB falls at the lower end of natural amplitude range of other avian species: song sparrows (*Melospiza melodia*) sing at an amplitude of 55dB to 85dB (Anderson et al., 2008), blackbirds (*Turdus merula*) at 74 to 79dB at a distance of 2 m (Dabelsteen, 1981), and chaffinches (*Fringilla coelebs*) at 78 to 87dB (Brumm & Ritschard, 2011). American crows were found to respond to distress call playback at a distance of up to 275m (Gorenzel et al., 2002), where the amplitude would have dropped from 96dB to approximately 48dB, so they should be able to perceive the stimuli at 65dB.

The crows' behavior was observed from a 10 – 15 m distance and field notes of observations were made with a Dictaphone, an established alternative to video recording behaviors (Kappeler, 2022) that allows notetaking of observations without needing to look away from the subject. Due to the distance between the experimenter and the crows, and the voice notes being made quietly, the crows should not have been able to hear the spoken notes. Therefore, this should not have interfered with the playback of speech stimuli. Video recordings were initially considered but would not have been feasible as head orientation (e.g., looking towards speaker) of the crows would be poorly visible when filmed from afar, and angles had to be changed quickly depending on the movement of the crows approaching the site, as well to avoid passers-by blocking the view.

Procedure

The data collection in London was conducted between 24th December 2020 and 27th March 2021. Data in Milton Keynes was collected from 27th December 2021 until 9th April 2022. Data collection had to

be terminated earlier than planned in Milton Keynes, due to an early onset of the breeding season, to avoid disturbing breeding crows and to avoid sampling bias towards male and non-breeding crows. The winter months were chosen for data collection because visitor rates to public parks are lowest from October to April (Hitchcock et al., 2008), reducing the risk of passers-by disturbing the experiment. It should, however, be noted that the London experiment coincided with lockdown measures in response to the COVID-19 pandemic, and visitation rates to parks may have been higher than in previous winters. The target sample size of 60 trials was determined *a priori* in G*Power, based on a One-Way ANOVA with three groups, an estimated effect size (Cohen's f) of 0.3, and minimum power of 0.8 (Faul et al., 2007). 60 trials in London and 54 trials in Milton Keynes (23 at the urban site, 31 at the suburban site) were included in the analysis.

The experiment began approximately 5 to 10 s after a crow had approached the bait (standing within 1m of it, either eating it, walking around it, or standing steadily, but not walking away). Focal anti-predator behaviors were vigilance and flight. Vigilance was measured by the crow looking towards the speaker (beak direction used as proxy for looking direction, so that this can be seen clearly in the field), as head position is commonly used as a proxy for vigilance in avian species (Mettke-Hofmann, 2022; Zhou et al., 2019). Flight was measured in the crow either, hopping, walking, or flying away from the speaker. These behaviors were chosen because they are anti-predator behaviors to varying intensities (Irigoin-Lovera et al., 2019), ranging from attention without movement to flight and abandonment of the site. Resumption of foraging was recorded as the end of vigilance and flight.

Playback was initiated when the crow was looking in any direction (based on beak direction) except towards the experimenter or the audio speaker, so that the behavior "looking towards the speaker" could be clearly observed. When multiple crows arrived at the same time, only one focal crow was observed (chosen randomly before playback onset to avoid bias). Flying away was considered a termination to the trial, as the crow would be out of earshot of the playback. Each voice note of the observation included the start of the playback (to calculate the time between onset and displayed behavior) as it happened, the focal behaviors as they happened, the group size and interaction between individuals if multiple birds were present, and the time at the end of the trial.

Analysis

Voice recordings were analyzed in Praat v.6.1.16 (Boersma & Weenink, 2020) where word onset is clearly visualized in the spectrogram. Focal behaviors were counted as response to the playback when they were noted on the voice recording within 2.5 s of playback onset. This accounts for a 2 s response latency by the crows (set following pilot trials as the crows responded immediately, and to ensure that responses recorded were triggered by the playback rather than co-occurring with the playback), and an additional 0.5s to account for the delay between me observing a behavior and verbalizing it on the voice note (determined prior to the start of data collection based on voice recording test runs).

All focal behaviors were coded as either presence (1) or absence (0). When two focal behaviors occurred, the behavior with higher urgency of the two was coded (looking < walking away < flying away). For instance, if a crow would first look towards the speaker and then flew away within the predefined playback onset period, the trial response would be recorded as "flying away". Length of vigilance was coded as seconds after onset when foraging was resumed after the last response behavior was recorded. If foraging was not resumed at all, 30 s was used as the longest possible time (maximum duration of trial). Interactions between heterospecifics (e.g., a crow and a magpie fighting over the food) and conspecifics were coded on three levels: no interaction (1), some interaction (2), and physical aggression (3).

The data was analyzed with a generalized linear model using the 'lme4' package in R (Bates et al., 2015; R Core Team, 2020) and were visualized in R (R Core Team, 2020) with the package ggplot2 (Wickham, 2016). I initially considered a generalized linear mixed model to account for pseudoreplication and differences between sites, but both lead to singular fit, and I used linear models instead. I analyzed the presence and absence of any behaviors per stimulus group (English, Vietnamese, control), each response individually (i.e., dropped all but one behavior), as well as walking and flying grouped into a "flight"

response as response ~ stimuli, with a binomial distribution. Next, I conducted the same analysis with English and Vietnamese grouped into speech.

The response intensity was analyzed with the same categories and analysis used by Irigoin-Lovera et al. (2019) for Peruvian guano birds (excluding wing flapping, which was not recorded here and not displayed by the crows): 0 for no reaction, 1 for head / beak pointing towards the speaker, 2 for walking away, and 3 for flying away. The intensity was analyzed as intensity ~ stimuli, with a Poisson distribution. I also calculated an intensity score for each study site as the average level across all observations at that site.

To understand whether the presence of conspecifics influenced the crows' responses, I compared the total count of responses for each stimulus type between trials in which conspecifics were present with those where the focal crow was alone (e.g., the total count of any response for "Group + Pigeon" compared to the total count of any response for "Alone + Pigeon", compared to the count for "Group + Vietnamese" and so on), using response ~ group stimuli with a Poisson distribution. Whether the response latency differed significantly between stimuli groups was tested with latency ~ stimuli, and gaussian distribution. Finally, I calculated the $_{log}$odds ratio (see formula 1 and 2) following Borenstein et al. (2009) for a response to speech, and a response to language. Log transformed odds ratios are used throughout, as Borenstein argues this transformation "is needed to maintain symmetry in the analysis" when comparing groups that differ in their weight.

$$1)\ OddsRatio = \frac{Response\ in\ Speech\ Trials * Non - Response\ in\ Non - Speech\ Trials}{Response\ in\ Non - Speech\ Trials * Non - Response\ in\ Speech\ Trials}$$

$$2)\ LogOddsRatio = ln\ (OddsRatio)$$

Results

London

At the experimental site in London, Vietnamese triggered a response in 13 out of 20 trials, English in nine out of 20 trials, and parakeet vocalizations in four out of 20 trials (Figure 1). The $_{log}$odds ratio for a response to the speech trials compared to the non-speech trials was 1.54, and 0.81 for Vietnamese compared to English.

For presence and absence of responses, there was a significant difference between English and the Parakeet control when walking and flying were grouped into a joined flight behavior (p = .02, z = -2.26, S E = 1.12). There was no significant effect of the stimuli variable for the presence of a flight behavior for Vietnamese and English (p = .74, z = -0.32, SE = 0.65), any response (Vietnamese: p = .2, z = 1.26, SE = 0.64; Parakeet: p = .09, z = -1.65, SE = 0.71), for looking on its own (Vietnamese: p = .06, z = 1.84, SE = 1.13; Parakeet: p = .31, z = 1, SE = 1.2), for walking on its own (Vietnamese: p = .43, z = -0.78, SE = 0.81; Parakeet: p = .99, z = 0, SE = 2404.67), or for flying away on its own (Vietnamese: p = .67, z = 0.41, SE = 0.83; Parakeet: p = .31, z = 1, SE = 1.2).

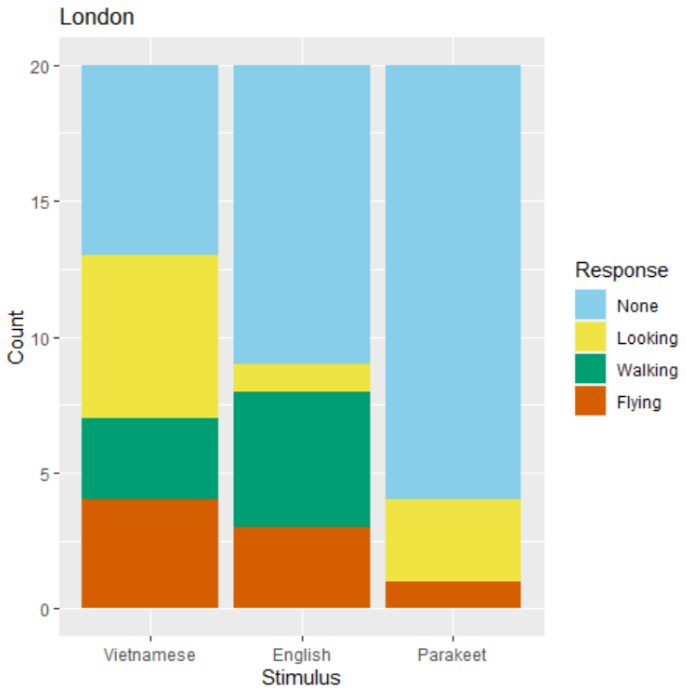
For presence and absence of responses by speech compared to control sounds, any response differed significantly between speech and no speech (p = .013, z = 2.46, SE = 0.64), as well as the combined flight response (p = .02, z = 2.26, SE = 1). There was no significant difference between speech and non-speech trials for looking on its own (p = .8, z = 0.24, SE = 0.75), walking away on its own (p = .99, z = 0, SE = 2404.67), or flying away on its own (p = .2, z = 1.25, SE = 1.1)

When responses are weighted based on intensity, there is a significant difference between English and Parakeet trials (p = .009, z = -2.58, SE = 0.46) but not between English and Vietnamese trials (p = .54, z = 0.6, SE = 0.3). The intensity score was 0.83 (95% CI [0.59, 1.06]).

The mean time lag between playback onset and first recorded behavior was 1.9 s for Vietnamese (SD = 1.7), 3.7 seconds for English (SD = 7.1), and 1.5 s for parakeet vocalizations (SD = 0.6, Figure 2), but differences were not significant for English compared to Vietnamese (p = .96, t = 0.04, S.E.=0.58), and not significant for English compared to the control (p = .68, t = -0.4, S.E.=0.85).

Figure 1

Response Types by Stimulus



Note. Number of each response type (looking towards speaker, walking away, flying away) for the 20 trials of each stimuli group (Vietnamese, English, Parakeet vocalizations).

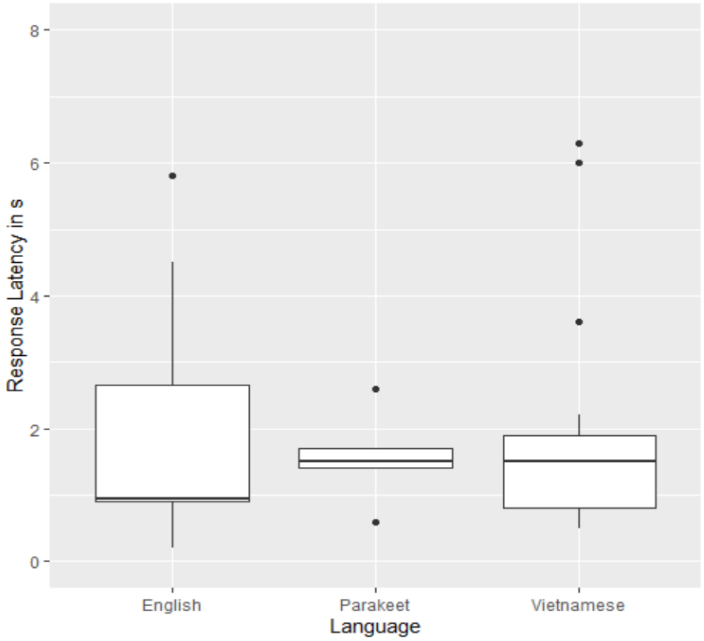
In trials with either looking or walking away behavior, foraging resumed on average 19 seconds after Vietnamese onset (10 trials, *SD* = 12), 20 seconds after English onset (6 trials, *SD* = 11), and 9.4 seconds after parakeet vocalizations onset (4 trials, *SD* = 6.1). As the latency to resumed foraging is only available for a smaller subset of trials (those where the crows did not leave the site and resumed foraging during the 30s trial), it was not included in the model analysis.

No responses were recorded in the last three trials with Vietnamese, the last four English trials, and the last five parakeet trials. All observation of crows flying away were made in the first 30 trials (first 10 for each stimuli group), whereas seven out of eight recordings of walking away were made in the last 30 trials (last 10 for each stimuli group). Looking towards the speaker was recorded throughout data collection (Figure 3). Two individuals began following me to the sites from the 30th trial (10th trial per stimuli group), as soon as I walked onto the Kenwood meadow.

Group sizes per trial were between one and 10 crows (M = 2.13, *SD* = 1.62). The presence of conspecifics did not have any significant effects (Alone + Parakeet: p = 1.0, z = 0, SE = 0.7; Alone + Vietnamese: p = .28, z = 1.07, SE = 0.81; Group + English: p = .29, z = 1.05, SE = 0.8; Group + Parakeet: p = .75, z = -0.31, SE = 0.91; Group + Vietnamese: p = .19, z = 1.28, SE = 0.8).

Figure 2

Response Latency by Stimulus



Note. Boxplots showing the response latency between stimuli onset and first recorded behavior for each stimuli group, only including trials where a behavior was recorded at any point during the playback (English = 14, Parakeet = 5, Vietnamese = 17).

Figure 3

Temporal Change of Responses

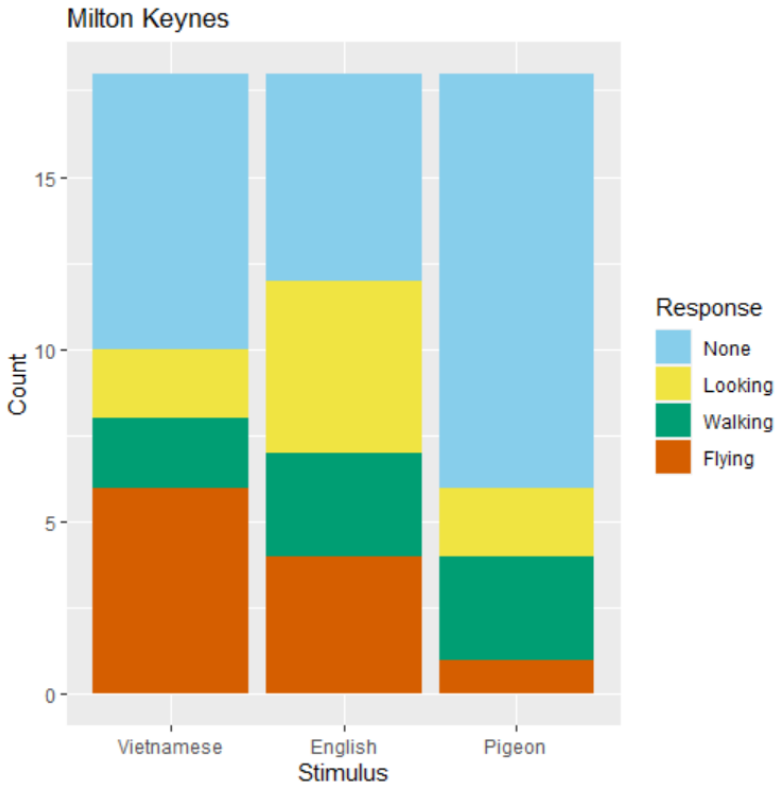
Note. Response towards the playback over time. Speech trials are shown as triangles, control trials as circles. Flying away is shown in purple, walking away in green, looking towards the speaker in orange, and trials with no responses are shown in blue.

Milton Keynes

Taking the two Milton Keynes sites together, Vietnamese triggered a response in 10 out of 18 trials, English in 12 out of 18 trials and pigeon vocalizations in six out of 18 trials (Figure 4), with a $_{log}$odds ratio for speech of 1.14 and a $_{log}$odds ratio for foreign language of -0.46 (Table 1).

Figure 4

Response Types By Stimuli



Note. Number of each response type (looking towards speaker, walking away, flying away) for the 18 trials of each stimuli group (Vietnamese, English, Pigeon vocalizations) pooled from both Milton Keynes sites.

Table 1

$_{log}$Odds ratios in response to speech and intensity scores across sites, calculated with formulas 1 and 2

Site	$_{log}$Odds Ratio Speech	Intensity Score
London	1.54	0.83
Milton Keynes Combined	1.14	1.07
Milton Keynes Suburban	1.63	0.83
Milton Keynes Urban	0.18	1.39

For presence and absence of responses, there was a significant effect of the stimuli variable for the presence of any response for English compared to the Pigeon control (p = .049, z = -1.96, SE = 0.7). There was no significant effect for any response for English compared to Vietnamese (p = 0.49, z = -0.68, SE = 0.68), for looking on its own (both: p = 0.21, z = -1.22, SE = 0.91), for walking on its own (Vietnamese: p = .63, z = -0.47, SE = 0.98; Pigeon: p = 1.0, z = 0, SE = 0.89), for flying on its own (Vietnamese: p = .45, z

= 0.74, SE = 0.75; Pigeon: p = .17, z = -1.34, SE = 1.17), or for walking and flying grouped into a joint flight response (Vietnamese: p = .73, z = 0.33, SE = 0.67; Pigeon: p = .28, z = -1.07, SE = 0.74).

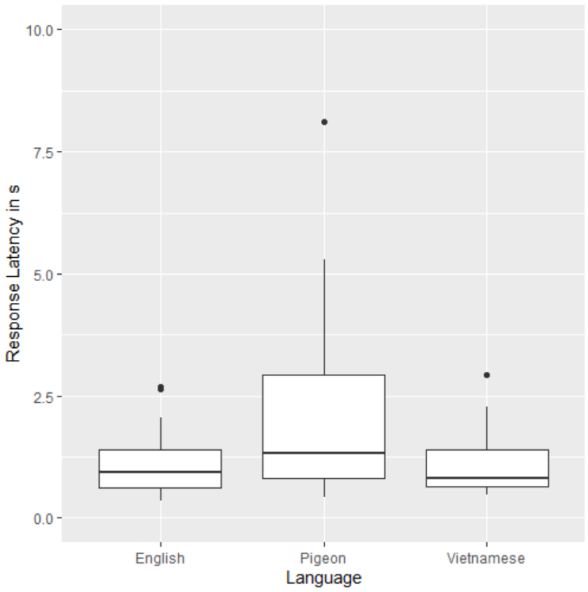
For presence and absence of responses by speech compared to control sounds, there was no significant effect for the presence of any response (p = .058, z = 1.89, SE = 0.6), looking on its own (p = .44, z = 0.76, SE = 0.86), walking on its own (p = .78, z = -0.27, SE = 0.79), flying away on its own (p = .08, z = 1.71, SE = 1.09), or walking and flying grouped into a joint flight response (p = .16, z = 1.38, SE = 0.66).

When responses are weighted based on intensity, there is a significant difference between English and Parakeet trials (p = 0.04, z = -2.01, SE = 0.36) but not between English and Vietnamese trials (p = .88, z = 0.14, SE = 0.29). The intensity score was 1.07 (95% CI [0.79, 1.34], Table 1).

The mean time lag between playback onset and first recorded behavior was 2.6 seconds for Vietnamese (95% CI [0.17, 5.02]), 1.2 seconds for English (95% CI [0.82, 1.57]), and 2.4 seconds for parakeet vocalizations (95% CI [1.14, 3.65], Figure 5). Response latency differences were not significantly different between English and Vietnamese (p = .27, t = 1.11, SE = 1.27), or between English and the control (p = .37, t = 0.9, SE = 1.33).

Figure 5

Response Latency by Stimuli



Note. Boxplots showing the response latency between stimuli onset and first recorded behavior for each stimuli group, only including trials where a behavior was recorded at any point during the playback (English = 14, Pigeon = 11, Vietnamese = 13).

In trials with either looking or walking away behavior, foraging was resumed 72% of the time. The average time from playback onset to resumed foraging was 15.2 seconds after Vietnamese onset (95% CI [7.36, 23.03]), 18.2 seconds after English onset (95% CI [11.03, 25.36]), and 8.5 seconds after pigeon vocalizations onset (95% CI [0.3, 16.69]).

At the suburban site, 41.9% of trials triggered a response, compared to 65.2% of trials at the urban site. The $_{log}$odds ratio for a speech response at the suburban site was 1.63, and 0.18 at the urban site. Foreign language $_{log}$odds ratios were 0.22 at the suburban site and -0.82 at the urban site (Table 1).

Looking only at trials conducted in suburban Milton Keynes (given the difference in $_{log}$odds ratios between suburban and urban Milton Keynes) for presence and absence of responses, there was a significant effect of the stimuli variable for the presence of any response for English compared to the Pigeon control

(p = .03, z = -2.1, SE = 1.2). There was no significant effect for any response for English compared to Vietnamese (p = .8, z = -0.25, SE = 0.88), for looking on its own (Vietnamese: p = .24, z = -1.15, SE = 1.25; Pigeon: p = .99, z = 0, SE = 3400.71), for walking on its own (Vietnamese: p = .91, z = -0.1, SE = 1.11; Pigeon: p = .53, z = -0.61, SE = 1.31), for flying away on its own (Vietnamese: p =.33, z = 0.97, SE = 1.25; Pigeon: p = .99, z = 0.0, SE = 3400.71), or for walking and flying grouped into a joint flight behavior (Vietnamese: p = .46, z = 0.72, SE = 0.91; Pigeon: p = .28, z = -1.07, SE = 1.25).

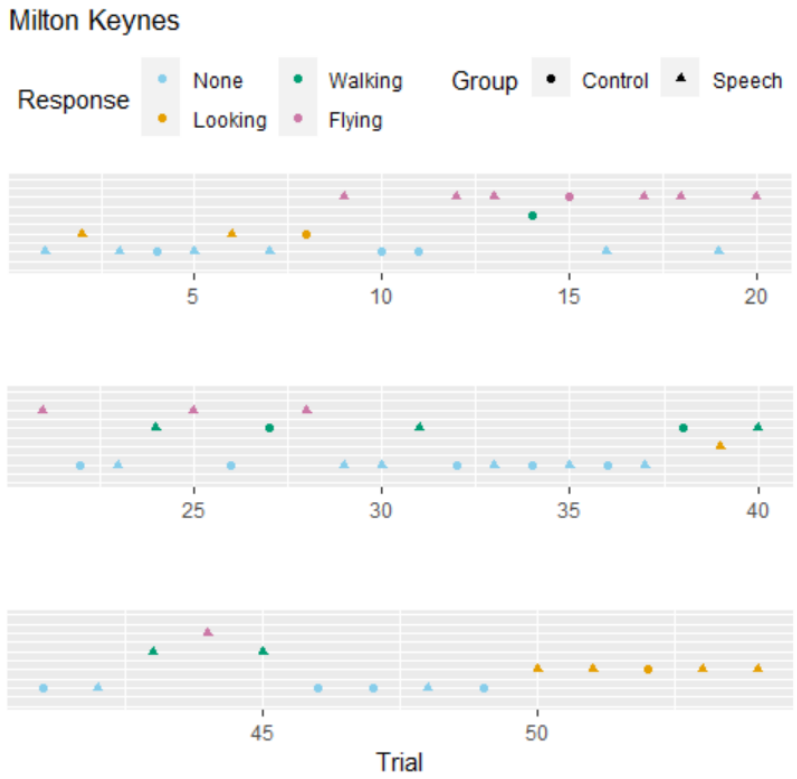
For speech responses in suburban Milton Keynes, there was a significant response for any behavior (p = .02, z = 2.17, SE = 1.14). There was no significant response for looking on its own (p = .99, z = 0, SE = 3400.72), walking on its own (p = .52, z = 0.63, SE = 1.19), flying on its own (p = .99, z = 0, SE = 3400.72), or walking and flying grouped into a joint flight response (p = .13, z = 1.49, SE = 1.14).

When suburban Milton Keynes responses are weighted based on intensity, there is a significant difference between English and Parakeet trials (p = .03, z = -2.07, SE = 0.77) but not between English and Vietnamese trials (p = .56, z = 0.58, SE = 0.41).

Additionally, there was no temporal pattern of decreased response rate over time, suggesting the crows at either site did not habituate to the playback (Figure 6).

Figure 6

Temporal Change of Responses



Note. Response towards the playback over time. Speech trials are shown as triangles, control trials as circles. Flying away is shown in purple, walking away in green, looking towards the speaker in orange, and trials with no responses are shown in blue.

Group sizes were between one and six crows (M = 2.07, *SD* = 1.06). When looking at stimulus type and presence of conspecifics, there was a significant difference for the presence of conspecifics during control trials (p = .02, z = -2.26, SE = 0.79), but not for any of the other stimuli – group types (Alone + Pigeon: p = .5, z = -0.66, SE = 3.53; Alone + Vietnamese: p = .3, z = -1.03, SE = 1.06; Group + English: p = .13, z = -1.49, SE = 0.61; Group + Vietnamese: p = .29, z = -1.05, SE = 0.48).

Discussion

To investigate response to human speech, carrion crows were exposed to playback of two languages (local and foreign) and nonhuman animal vocalizations. In both cities, the crows showed a significant difference in their response intensity to English compared to the control bird vocalizations. In London, the crows also responded significantly more with flight behaviors (walking and flying away) to English compared to the parakeet control. When looking at speech compared to non-speech trials, the London crows responded significantly more overall and responded significantly more often with flight behaviors. In suburban Milton Keynes, the crows responded significantly more to English compared to pigeon recordings and more to speech compared to non-speech overall. When Milton Keynes sites were pooled together, the crows still responded significantly more often to English than to the pigeon recordings.

Neither London nor Milton Keynes crows responded differently to English than to Vietnamese, though they did respond insignificantly quicker to English than to either Vietnamese or control trials, possibly due to existing familiarity with the local language. The equal response to the familiar and the unfamiliar language is unlike the differential pattern observed for large-billed crows (Schalz & Izawa, 2020), and the absence of a novelty response to the unfamiliar language is unlike the response observed in domestic dogs (Mallikarjun et al., 2022). This suggests at least three possible explanations:

1. Carrion crows are unable to perceive the differences between languages and only perceive the superset of speech. If this is the case, they would also be unable to learn the differences between English and Vietnamese if trained.
2. Carrion crows perceive the differences between English and Vietnamese, but that distinction is not relevant to them as both groups of acoustic cues are speech and therefore equally indicate human presence. This explanation presupposes that they group English and Vietnamese together by their shared linguistic cues while disregarding those cues that differ between them. If this is the case, they would be able to learn the differences between English and Vietnamese if there was a sufficient reward or risk associated with one language but not the other.
3. Carrion crows would be able perceive Vietnamese and English as different languages but would only perceive the difference as relevant in a monolingual habitat. Due to the high degree of linguistic variety in London and some linguistic variety in Milton Keynes, the crows in these two cities have habituated to exposure to multiple languages and their speech template is broader than that of crows in less multilingual areas. This explanation assumes overgeneralization due to exposure to many, possibly contradictory cues gathered from multiple languages.

Future experiments are needed to investigate whether any of these three possible explanations should be considered further, either in a field experiment with wild individuals or in an aviary with captive crows. It is also important to note that since individual crows could not be identified, it is unclear whether any individual participated in the experiment several times, and if so, how many times. Especially towards the end of data collection, when I noticed two crows flying ahead of me to the usual site, it seemed likely that some individuals had learned the association between the experiment and the food provided as bait, thus participating several times (though these two individuals did still respond to the playback on the day they flew ahead to the site, suggesting they associated me with the food, but not with the playback coming from the shrubs). This risk of pseudo replication, intensified by some individuals potentially recognizing me, needs to be taken into consideration when evaluating the results presented here.

The response to speech suggests that this behavior is relevant to predator avoidance, specifically to reduce the time spent in close proximity to humans, who can sometimes pose a threat to crows. American crows (*Corvus brachyrhynchos*) in Seattle and hooded crows (*Corvus cornix*) in Berlin show an increased flight initiation distance when local discouragement levels are high (Clucas & Marzluff, 2012). In areas of Japan where they are being shot, carrion crows and large-billed crows show a greater alert distance and flight initiation distance than their conspecifics in areas where they are being trapped instead (Fujioka, 2020). This is unlike findings in wild jackdaws, which were less likely to respond to human voices if they

had been exposed to high disturbance in the past, and who did not respond differently based on familiarity with the speaker, or threat level (McIvor et al., 2022). Some London residents report feeding wild crows, while others try to scare crows away or chase them away (Schalz, 2021). Eavesdropping on speech may provide useful information in addition to visual scanning, to improve monitoring for human presence and thereby avoid close proximity to humans who might chase crows.

Taking into account the higher response to Dutch than to Japanese shown by the large-billed crows (Schalz & Izawa, 2020), the present findings suggest that there is either a considerable difference in speech perception abilities between the two species, or that the large-billed crows had such monolingual exposure to Japanese that both non-Japanese phonemes and non-Japanese prosody were not sufficiently encountered. Considering that 97% of the population in Tokyo identifies as Japanese (Statistics Bureau of Japan, 2017), spoken languages other than Japanese will rarely be encountered aside from tourist locations. While 89% of Milton Keynes households use English as their main language (Office for National Statistics, 2011), that still leaves 11% of residents speaking other languages, which may include other East Asian languages like Vietnamese, as well as other languages with syllable-timed prosody (such as French or Spanish). Tamarin monkeys (Ramus et al., 2000) and rats (Toro et al., 2003) use prosody to discriminate between Dutch and Japanese, and it would be worth further investigation whether carrion crows are able to perceive this cue.

It should be noted that the crows occasionally responded with either vigilance or flight behaviors after the 2 s threshold. For instance, some individuals did not move during the pre-defined onset period but still flew away a few seconds later. Individual reactions will also be influenced by the individual's previous experience with humans; some will have had negative experiences and might flee immediately when hearing human speech, while some appeared unsure of how to react and slightly delay their flight, and others do not take flight at all. Foraging also resumed later during speech playback than it did during parakeet playback, suggesting prolonged vigilance after the initial reaction during playback onset.

Additionally, it is worth keeping in mind that when the crows did respond to the playback within the pre-defined playback onset timeframe, they did so within the first few seconds of the 30 seconds-long playlist. This means they responded to only the very first stimulus in the playlist. Their average response times to the speech playlists was between one and four seconds, which corresponds to the first half of the first sentence played in each trial, approximately covering between two and six words. This may limit the extent of prosodic information available, such as prosodic phrases and global prosodic patterns spanning the entire sentence (Carlson, 2009; Frazier et al., 2006). However, information such as word-level prosody (Arciuli & Slowiaczek, 2007) and phonemes would still be available in this time frame.

Over the course of the experiment, London crows responded to speech with decreasing intensity: initially by flying away, then by walking away, and eventually they stopped responding entirely. Milton Keynes crows did not seem to habituate to the playback, although suburban Milton Keynes crows did approach me and the experimental site quicker as the experiment progressed, possibly learning the association between the experimenter, the speaker, and the food. This is similar to the two crows flying ahead of me to the playback site in London. It is unclear whether this recognition affected their response to the playback.

There was a considerable difference in responses between the two Milton Keynes sites, with the crows at the suburban site having a speech response $_{\log}$odds ratio slightly above that of London crows, while the $_{\log}$odds ratio for the urban Milton Keynes crows was close to 0. Intensity scores and the count of response presence were also considerably higher for urban than suburban Milton Keynes crows, suggesting this group was overall more vigilant to any stimulus than the suburban one. Most suburban trials did not trigger any response, while most urban trials did. Crows at the suburban site showed their first response later after playback onset than individuals at the urban site, and resumed foraging more often while playback was still ongoing. Suburban crows were also twice as likely to walk towards the speaker during playback, with some individuals standing right next to it during playback and approaching rates being equal for all stimuli. Higher response rates at the urban compared to the suburban site is contrary to expectations, as urban birds are usually bolder (Gravolin et al., 2014; Vines & Lill, 2015), are more habituated to humans (Stansell et al., 2022), and habituate faster to human presence (Vincze et al., 2016) than their rural conspecifics.

Site-specific cost-benefit differences may drive this difference in behavior. The urban site has few foraging opportunities (the two small meadows and potential dropped food in front of the station) but a high rate of human presence (personal observations), suggesting a low foraging pay-off paired with a high predation risk. Given that an individual should abandon a patch when the (perceived) cost of remaining outweighs the cost of fleeing (Ydenberg & Dill, 1986), any unexpected sound may be enough to trigger flight on a patch with lower foraging potential and high predation risk. The suburban site on the other hand has few human visitors (personal observations) and the football field is a large, open grass area well suited for foraging. Flight should be costlier in a patch like this in terms of lost foraging opportunities, and so should be triggered less often. Likewise, responses should be less sensitive to ambiguous cues on high-quality patches like this. The large-billed crows in Tokyo (Schalz & Izawa, 2020) had been placed into an outdoor aviary (one individual at a time), which prevented them from taking flight, and this restriction may have also influenced their behavior towards the playback. The effect of patch quality on vigilance and flight responses towards speech would be worth further investigation in the future.

Conclusion

Overall, both London and Milton Keynes crows respond more often and with higher intensity to speech playback compared to bird control playback. These findings show that crows attend to human speech in the wild, and not only in captivity. Despite a considerable difference in local human population density, suburban Milton Keynes crows responded to speech with the same intensity as crows in London. Additionally, Milton Keynes crows did not respond differently to English than to Vietnamese despite a high share of English-speaking households, suggesting that carrion crows are either unable to perceive the difference between the two languages, that they do not perceive a meaningful difference between them (as all languages indicate human presence, and all humans could be dangerous), or that they would only respond differently in a monolingual environment.

Acknowledgements

I am very grateful to Tom Dickins and Martijn Timmermans for their supervision, patience, and critical review. I also thank Kaeli Swift for her invaluable advice to use Doritos as bait, and all crows who participated in the experiment.

Data Accessibility: R scripts and data used in this experiment can be accessed freely via: https://doi.org/10.6084/m9.figshare.c.6295581.v2

Conflicts of interest: The author declares no conflict of interest.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Aldine Pub. Co. https://books.google.co.uk/books/about/Elements_of_general_phonetics.html?id=SvVYAAAAMAAJ&redir_esc=y

Anderson, R. C., Searcy, W. A., Peters, S., & Nowicki, S. (2008). Soft Song in Song Sparrows: Acoustic Structure and Implications for Signal Function. *Ethology*, *114*(7), 662–676. https://doi.org/10.1111/J.1439-0310.2008.01518.X

Arciuli, J., & Slowiaczek, L. M. (2007). The where and when of linguistic word-level prosody. *Neuropsychologia*, *45*(11), 2638–2642. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2007.03.010

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

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/JSS.V067.I01

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). Effect Sizes Based on Binary Data (2 2Tables). In *Introduction to Meta-Analysis* (2nd ed.).

Bose, S., & Skinner, C. (2009). *Westminster Open Spaces Noise Study 2008 For Westminster City Council*.

Brumm, H., & Ritschard, M. (2011). Song amplitude affects territorial aggression of male receivers in chaffinches. *Behavioral Ecology*, *22*(2), 310–316. https://doi.org/10.1093/BEHECO/ARQ205

Carlson, K. (2009). How Prosody Influences Sentence Comprehension. *Language and Linguistics Compass*, *3*(5), 1188–1200. https://doi.org/10.1111/J.1749-818X.2009.00150.X

Clucas, B., & Marzluff, J. M. (2012). Attitudes and actions toward birds in urban areas: Human cultural differences influence bird behavior. *The Auk*, *129*(1), 8–16. https://doi.org/10.1525/auk.2011.11121

Cuaya, L. V., Hernández-Pérez, R., Boros, M., Deme, A., & Andics, A. (2021). Speech naturalness detection and language representation in the dog brain. *NeuroImage*, 118811. https://doi.org/10.1016/J.NEUROIMAGE.2021.118811

Dabelsteen, T. (1981). The Sound Pressure Level in the Dawn Song of the Blackbird Tardus merula and a Method for Adjusting the Level in Experimental Song to the Level in Natural Song. *Zeitschrift Für Tierpsychologie*, *56*(2), 137–149. https://doi.org/10.1111/J.1439-0310.1981.TB01292.X

Di Giovanni, J., Fawcett, T. W., Templeton, C. N., Raghav, S., & Boogert, N. J. (2022). Urban gulls show similar thermographic and behavioral responses to human shouting and conspecific alarm calls. *Frontiers in Ecology and Evolution*, *0*, 858. https://doi.org/10.3389/FEVO.2022.891985

Dutour, M., Walsh, S. L., Speechley, E. M., & Ridley, A. R. (2021). Female Western Australian magpies discriminate between familiar and unfamiliar human voices. *Ethology*, *127*(11), 979–985. https://doi.org/10.1111/ETH.13218

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3:A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPower3-BRM-Paper.pdf

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249. https://doi.org/10.1016/J.TICS.2006.04.002

Fujioka, M. (2020). Alert and Flight Initiation Distances of Crows in Relation to the Culling Method, Shooting or Trapping. *Https://Doi.Org/10.2326/Osj.19.125*, *19*(2), 125–134. https://doi.org/10.2326/OSJ.19.125

Gorenzel, W. . P., Salmon, T. P., Pearson, A. , B., & Ryan, S. R. (2002). Sound levels of broadcast calls and responses by American crows. *Proceedings of the Vertebrate Pest Conference*, *20*(20). https://doi.org/10.5070/v420110017

Gravolin, I., Key, M., & Lill, A. (2014). Boldness of Urban Australian Magpies and Local Traffic Volume. *Avian Biology Research*, *7*(4), 244–250. https://doi.org/10.3184/175815514X14151981691872

Hirst, D., & Di Cristo, A. (1998). A survey of intonation systems. In *Intonation Systems: A Survey of Twenty Languages* (pp. 1–44). Cambridge University Press.

Hitchcock, M., Curson, T., & Parravicini, P. (2008). *VISITORS TO THE ROYAL PARKS: RESULTS OF STEADY STATE COUNT (AUGUST 2007-JULY 2008)*.

Irigoin-Lovera, C., Luna, D. M., Acosta, D. A., & Zavalaga, C. B. (2019). Response of colonial Peruvian guano birds to flying UAVs: Effects and feasibility for implementing new population monitoring methods. *PeerJ*, *7*, e8129. https://doi.org/10.7717/peerj.8129

Kappeler, P. M. (2022). *Animal Behaviour : An Evolutionary Perspective* (5th ed.). Springer. https://blackwells.co.uk/bookshop/product/9783030828783?gC=5a105e8b&gclid=CjwKCAjwrfCRBhAXEiwAnkmKmfn7hN-Fomo6_BlqvjmyOLJ4JZjrtVjgcj1csuJJ0dXE8KT0bN5KXRoCv4UQAvD_BwE

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics* (7th ed.). Wadsworth Publishing.

Luong, H. T., & Vu, H. Q. (2016). *A non-expert Kaldi recipe for Vietnamese Speech Recognition System*. http://ailab.hcmus.edu.vn/vivos.

Magrath, R. D., Haff, T. M., Fallow, P. M., & Radford, A. N. (2015). Eavesdropping on heterospecific alarm calls: from mechanisms to consequences. *Biological Reviews*, *90*(2), 560–586. https://doi.org/10.1111/BRV.12122

Mallikarjun, A., Shroads, E., & Newman, R. S. (2022). Language preference in the domestic dog (*Canis familiaris*). *Animal Cognition*, *26*, 451–463. https://doi.org/10.1007/S10071-022-01683-9

McIvor, G. E., Lee, V. E., & Thornton, A. (2022). Nesting jackdaws' responses to human voices vary with local disturbance levels and the gender of the speaker. *Animal Behaviour*, *192*, 119–132. https://doi.org/10.1016/J.ANBEHAV.2022.08.006

Mettke-Hofmann, C. (2022). Is vigilance a personality trait? Plasticity is key alongside some contextual consistency. *PLOS ONE*, *17*(12), e0279066. https://doi.org/10.1371/JOURNAL.PONE.0279066

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766. https://psycnet.apa.org/buy/1998-02354-006

Nguyễn, Đ. L. (1970). *Vietnamese Pronunciation*. University of Hawaii Press. https://books.google.co.uk/books/about/Vietnamese_pronunciation.html?id=CL6CAAAAIAAJ&redir_esc=y

Office for National Statistics. (2011). *Main Language Spoken at Home (Census), Borough*. https://data.london.gov.uk/dataset/main-language-spoken-at-home-borough

Office for National Statistics. (2022). *Population and household estimates, England and Wales: Census 2021*. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationandhouseholdestimatesenglandandwalescensus2021

Pfitzinger, H. R., & Kaernbach, C. (2008). Amplitude and amplitude variation of emotional speech. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 1036–1039. https://doi.org/10.21437/INTERSPEECH.2008-322

Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Press. https://www.amazon.com/Intonation-American-English-Kenneth-Pike/dp/B002IW5FUI

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

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*(5464), 349–351. https://doi.org/10.1126/science.288.5464.349

Robinson, R. A., Leech, D. I., & Clark, J. A. (2020). *The Online Demography Report: bird ringing and nest recording in Britain & Ireland in 2019*. BTO, Thetford. http://www.bto.org/ringing-report

Schalz, S. (2021). Attitudes and Behaviours Towards Carrion Crows in London. *London Naturalist*. https://www.researchgate.net/publication/356727449_Attitudes_and_Behaviours_Towards_Carrion_Crows_in_London

Schalz, S., & Izawa, E.-I. (2020). Language Discrimination by Large-Billed Crows. In A. Ravignani, C. Barbieri, M. Martins, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, K. Mudd, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 13th International Conference (EvoLang13)*. https://doi.org/10.17617/2.3190925

Stansell, H. M., Blumstein, D. T., Yeh, P. J., & Nonacs, P. (2022). Individual variation in tolerance of human activity by urban Dark-eyed Juncos (Junco hyemalis). *Wilson Journal of Ornithology*, *134*(1), 43–51. https://doi.org/10.1676/21-00001

Statistics Bureau of Japan. (2017). *POPULATION AND HOUSEHOLDS OF JAPAN*. http://www.stat.go.jp/english/data/kokusei/2015/final_en/final_en.html

Suraci, J. P., Clinchy, M., Zanette, L. Y., & Wilmers, C. C. (2019). Fear of humans as apex predators has landscape-scale impacts from mountain lions to mice. *Ecology Letters*, *22*(10), 1578–1586. https://doi.org/10.1111/ele.13344

Tang, G. M. (2007). Cross-Linguistic Analysis of Vietnamese and English with Implications for Vietnamese Language Acquisition and Maintenance in the United States. *Journal of Southeast Asian American Education & Advancement*, *2*(1). https://doi.org/10.7771/2153-8999.1085

Toro, J. M., Trobalon, J. B., & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition*, *6*(2), 131–136. https://doi.org/10.1007/s10071-003-0172-0

Vincze, E., Papp, S., Preiszner, B., Seress, G., Bókony, V., & Liker, A. (2016). Habituation to human disturbance is faster in urban than rural house sparrows. *Behavioral Ecology*, *27*(5), 1304–1313. https://doi.org/10.1093/beheco/arw047

Vines, A., & Lill, A. (2015). Boldness and urban dwelling in little ravens. *Wildlife Research*, *42*(7), 590. https://doi.org/10.1071/WR14104

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

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Ydenberg, R. C., & Dill, L. M. (1986). The Economics of Fleeing from Predators. *Advances in the Study of Behavior*, *16*(C), 229–249. https://doi.org/10.1016/S0065-3454(08)60192-8

Zhou, Y., Radford, A. N., & Magrath, R. D. (2019). Why does noise reduce response to alarm calls? Experimental assessment of masking, distraction and greater vigilance in wild birds. *Functional Ecology*, *33*(7), 1280–1289. https://doi.org/10.1111/1365-2435.13333/SUPPINFO

Supplementary Materials

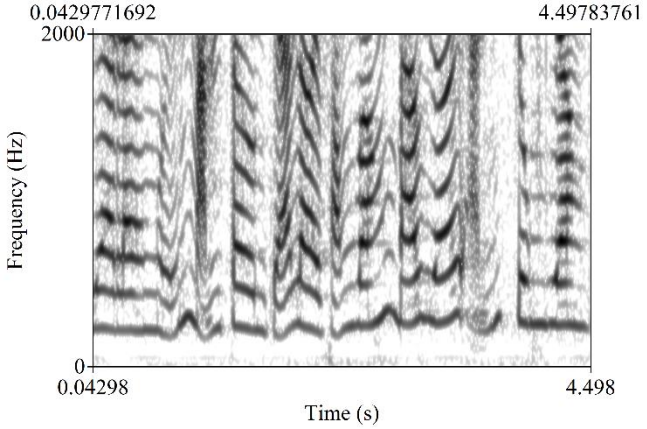
Parakeet and wood pigeon control recordings were the below recordings obtained from xeno-canto:

1. Parakeet stimuli sources
    a. Dick de Vos, XC561018
    b. Cedric Mroczko, XC534457 & XC534454
    c. Fernand Deroussen, XC528250

2. Wood Pigeon stimuli sources
    a. Manceau Lionel, XC653672
    b. Domagoj Tomičić, XC656200
    c. Mats Olsson, XC651508

Figure S1

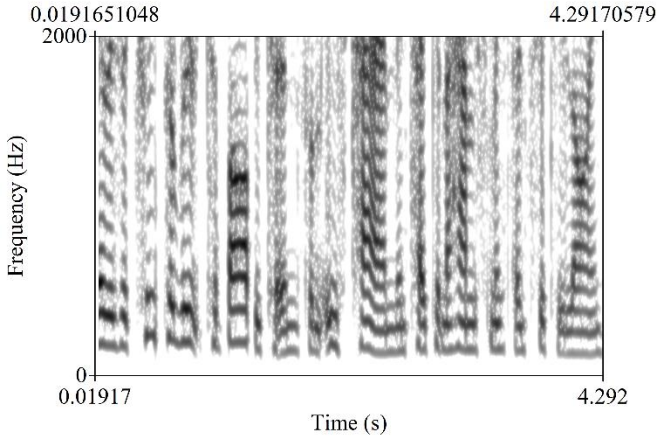
Spectrogram for one of the Vietnamese stimuli used in the experiment



Note. Created in Praat (Boersma & Weenink, 2020).

Figure S2

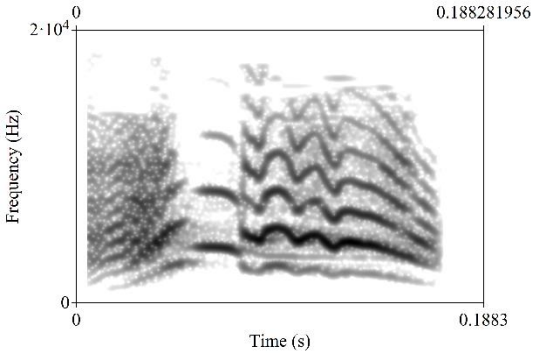
Spectrogram for one of the English stimuli used in the experiment



Note. Created in Praat (Boersma & Weenink, 2020).

Figure S3

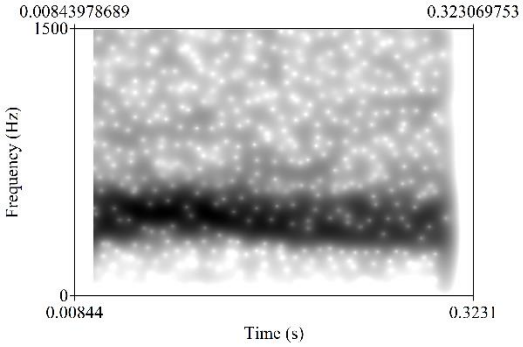
Spectrogram for one of the parakeet stimuli used in the experiment



Note. Created in Praat (Boersma & Weenink, 2020).

Figure S4

Spectrogram for one of the pigeon stimuli used in the experiment



Note. Created in Praat (Boersma & Weenink, 2020)