The Acoustic Discrimination and Recognition of Individual Heterospecifics by Humans:

Proximate and Ultimate Causes

A thesis submitted to Middlesex University in partial fulfilment of the requirements for the degree of Master of Science by Research

Sabrina Schalz

M00727483

Department of Psychology

Middlesex University London

May 2020

Abstract

Humans readily discriminate and recognize others by their voices, but it has not yet been studied whether this is also possible with the vocalizations of individual non-primates. My thesis addresses this question from both the proximate and ultimate perspective. In Chapter 1, I introduce the mechanism behind the individual discrimination of conspecific and heterospecifics by their vocalizations. Using a comparative approach, the physical properties of vocalizations and individual voices, their production and perception by species across the amniote clade are discussed. Three experiments are presented to examine whether humans can discriminate individual zebra finches by their songs with and without pitch contour, as well as individual large-billed crows by their calls. In Chapter 2, I focus on the functions associated with individual recognition through vocalizations and what fitness benefits this may entail, specifically in the context of reciprocal altruism between heterospecifics. A fourth experiment assesses the acoustic recognition of individual crows by humans in an altruistic exchange situation. In Chapter 3, I employ a theoretical approach reviewing the phylogenetic history of the focal species and their last common ancestor with a focus on their respective vocalizations. Taken together, these three sections provide further insights into the evolution of amniote vocalizations and the perception of identity across species.

Results obtained in Experiments 1 and 2 have been submitted for publication.

Acknowledgements

I am deeply grateful to my supervisor Tom Dickins, who has been an excellent mentor. Thank you for your comments on this thesis, for being so supportive, and for not losing your patience with me when I worry about nothing. I would also like to thank Tamar Keren-Portnoy and Fabia Franco for their help with the initial conceptualization of this thesis, Emma Ward for helpful suggestions during the registration review, as well as Barbara Caspers and Ei-Ichi Izawa for enabling me to record the zebra finch songs and crow calls.

I want to thank my father and stepmother for their unconditional love and support. No amount of words could adequately express my gratitude to both of you. I am also very grateful to my dearest friends Katharina and Jonas, who have suffered through endless monologues about my research for years and somehow still find it in their hearts to listen, encourage, and support me. Finally, I would like to thank the nameless crow who inspired Chapter 2.

Table of Contents

Chapter 1: Mechanism

1. Introduction

The faculty of language in a broad sense includes abilities and concepts related to vocalizations, some of which are shared with other non-human animals and others that appear unique to humans (Hauser, Chomsky, & Fitch, 2002). These unique abilities include vocal communication systems, a complex Theory of Mind, complex vocal learning, and supra-regularity. The shared foundation of speech consists of (amongst others) working memory, finite-state grammar, concepts, categories, as well as planning, transitive inference, emotional expression, vocal control, and the source-filter theory (Fitch, 2017). The source-filter theory constitutes the foundation of this chapter, as it describes the shared mechanism of tetrapod sound production and explains how vocalizations can transmit identity cues. In the past it has been argued that the diverse communication systems used by different species are “mutually incomprehensible” (Hauser et al., 2002). This hypothesis will be challenged in the second part of the introduction, as well as in the subsequent three experiments. The primary focus is on zebra finches, large-billed crows, and humans as they are the focal species in this thesis, but the vocalizations of other amniotes are taken into account as well.

1.1. The Source-Filter Theory

According to the source-filter theory, a sound (the source signal) is produced in the larynx or syrinx (Fitch, 1999; Taylor & Reby, 2010) and then filtered in the vocal tract (Chiba & Kajiyama 1941 as cited in Taylor, Charlton, & Reby, 2016; Fant, 1960; Titze, 1994). The anatomical structure of the larynx and the vocal tract will be discussed in detail in the next section. The source signal contains the fundamental frequency F0 that is perceived as pitch by the listener. F0 is the result of vibrations of the vocal folds and depends on their size, as well as the speed of the glottis opening and closing (Taylor & Reby, 2010; Titze, 1994). F0 is also the first harmonic of a complex sound, as harmonics are the multitudes of the fundamental frequency (Harrington & Cassidy, 1999). This source signal is then filtered in the vocal tract, an air-filled cavity (the pharynx and the mouth in humans), by amplifying some frequencies and decreasing others depending on its shape and resulting resonant properties. The resonant properties can be altered through movement of the tongue, jaw and mouth (Taylor & Reby, 2010; Titze, 1994). For instance, if you produce the vowel [a] followed by [i], you will notice that you move these components of the vocal tract, thereby changing the shape of your vocal tract and its resonant properties. Vocal flexibility differs between species and is generally more restricted in non-human animals (Fitch, 1994). The vibrating air within the vocal tract results in areas of enhanced frequency of the source signal, the formants (Fant 1960, as cited in Taylor et al., 2016; Taylor & Reby, 2010, see figures 1 and 2).

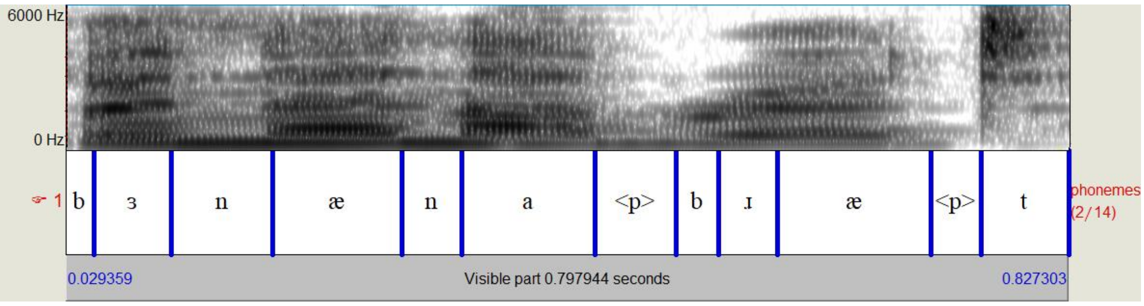


Figure 1: Exemplary spectrogram of the words “banana bread” (female voice, non-native), with a frequency window of 0Hz to 6000Hz created in Praat version 6.0.49 (Boersma & Weenink, 2019), annotated for reference following standard IPA annotation. The darkness of an area indicates how much energy it has, and particularly dark areas are formants (more energy due to amplification of the vocal tract). Vowels are characterised by their respective formants and these differences can be seen in this spectrogram when comparing the distinct phonemes. Particularly light areas are silence.

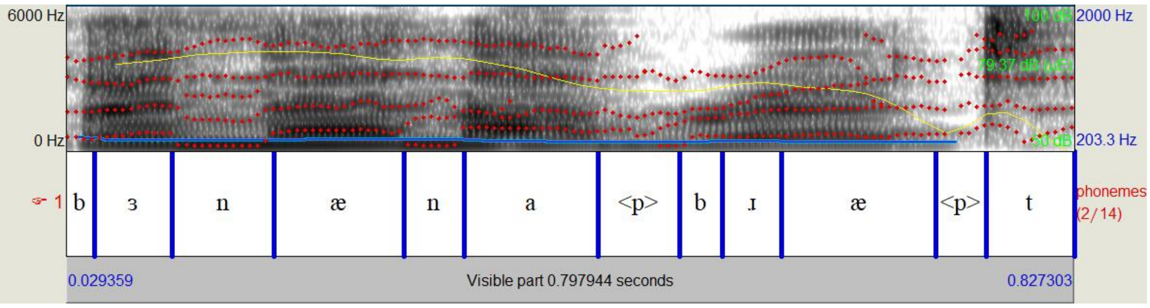


Figure 2: Exemplary spectrogram of the words “banana bread” (female voice, non-native) with a frequency window of 0Hz to 6000Hz created in Praat version 6.0.49 (Boersma & Weenink, 2019), annotated for reference following standard IPA annotation. A pitch line (blue; Praat measures pitch instead of F0), intensity line (yellow), and formants (red dots) are also shown. Pitch was set to 50Hz to 2000Hz, intensity was set to 50dB to 100dB, formants were set to 5 with a maximum frequency of 5000Hz. Mean pitch (203.3Hz) is indicated on the right in blue below the set maximum pitch. As explained above, the formant dots are shown on the darker areas in the spectrogram.

Formant dispersion, the frequency distance between two neighbouring formants (e.g. F1 and F2), conveys information about the body size of the vocalizer as they depend on the length of the vocal tract. This type of physical constraint often guarantees honesty of the signal to the receiver independently of the cost of sound production (Fitch, 1997), except for species in which acoustic size exaggeration evolved (Charlton & Reby, 2016). Men have been found to estimate the body size of an animal more accurately than women based on synthetic formant differences of the animal’s vocalizations, suggesting that this sensitivity is subject to sexual selection in humans as well (Charlton, Taylor, & Reby, 2013). Other types of information about the signal sender include sex, as males typically produce lower fundamental frequencies (Huber, Stathopoulos, Curione, Ash, & Johnson, 1999), while temporal cues such as rate indicate physical arousal, e.g. when facing a threat (Manser, 2001), physical condition (Pitcher, Briefer, Vannoni, & McElligott, 2014), or emotional state (Taylor, Reby, & McComb, 2009). The distribution of formants then leads to the formation of timbre, the sound quality of the source signal (Childers & Lee, 1991). Timbre is "the attribute that distinguishes sounds of equal pitch, loudness, location and duration" (Town & Bizley, 2013). Voice timbre is an important discrimination feature, one that musicians are significantly better at recognizing than non-musicians, together with tonal pitch differences (Chartrand & Belin, 2006; Pitt, 1994). Fundamental frequency range, frequency modulation, the spacing between formants, and bandwidth also play a role in individual discrimination.

These identity cues are used by multiple species, including Humboldt and Magellanic penguins (Favaro, Gamba, Gili, & Pessani, 2017) and meerkats (Townsend, Charlton, & Manser, 2014). Vocal signatures provide additional discrimination information through extraordinarily distinctive fundamental frequency contours (Taylor et al., 2016; Titze, 1994). Large-billed crows (*Corvus macrorhynchos*) have a signature voice system in which variation of their innate ka-calls between individuals is mostly introduced by characteristics of the fundamental frequency and temporal cues (Kondo, Izawa, & Watanabe, 2010). Other species with individually distinct vocalizations include the Ryukyu scops owl (*Otus elegans*; Takagi, 2020) or the Eastern grey wolves (*Canis lupus Lycaon*; Root-Gutteridge et al., 2014). Fundamental frequency as well as formant frequency also play a major role in the discriminability of human vocalizations (Baumann & Belin, 2010; Matsumoto, Hiki, Sone, & Nimura, 1973), although fundamental frequency appears to be the most important cue to determine a difference between two speakers, followed by F1 for female voices and the dispersion between F4 and F5 for male voices (Baumann & Belin, 2010). Individual differences in pitch contour remain stable across different types of vocalizations, such as speech, screams, roars, and pain cries (Pisanski, Raine, & Reby, 2020). Other discriminating factors include different phonation types (such as breathy, creaky, open or closed; Gordon & Ladefoged, 2001), prosodic structure of speech (Kim, 2019), sociocultural differences (such as pitch frequency on perceived femininity, see van Bezooijen, 1995) and sociolinguistic cues (such as dialects or accents, see Clopper & Pisoni, 2004).

The identity of signal senders is thus transmitted through passive cues mostly dependent on the physical properties of the vocal tract, specifically vocal tract filtering effects (Rendall, Owren, & Rodman, 1998) but as mentioned above, these cues may also be supported by individual signatures. A signature song is a vocalization acquired through vocal learning with a stereotyped pattern that differs inter-individually (Zann, 1997). Signature songs may compensate for insufficient voice cues in species with high-frequency vocalizations, an issue that may arise as formant dispersion is more accurate in low-frequency vocalizations (Fitch, 1997). The signature song of the zebra finch (*Taeniopygia guttata*), an Australian songbird (Elie & Theunissen, 2016), consists of multiple motifs that are arranged in a stereotyped order and separated by long intervals of at least 2 seconds. Motifs in turn consist of different elements that are separated by shorter intervals and preceded by introductory elements. Motifs include acoustic cues about the individual's identity, whereas introductory elements show little variance between males (Sossinka & Böhner, 1980). Their song consists of multiple acoustic components: Envelope cues span the entire song, whereas fine structure cues are limited to individual elements. Temporal cues include the rhythm of the song, the duration and arrangement of elements (Vernaleo & Dooling, 2011), whereas spectral cues refer to formants (Elie & Theunissen, 2016), fundamental frequency, harmonics (Lachlan, van Heijningen, Ter Haar, & ten Cate, 2016), and timbre (patterns of harmonic suppression; Williams, Cynx, & Nottebohm, 1989). Temporal fine structure alone is enough for zebra finches to discriminate syllables, although other acoustic cues are also relevant (Vernaleo & Dooling, 2011; Vernaleo, Dooling, & Leek, 2010). In addition to its relevance for syllable discrimination, fine structure also contains information about the individual's identity, its sex and the specific call type (Prior, Smith, Lawson, Ball, & Dooling, 2018).

Initially, the source filter theory was thought to only apply to a small number of species, but more recent research suggests that it applies to amniotes in general. Reber, Nishimura, Janisch, Robertson, and Fitch (2015) showed that the bellows produced by Chinese crocodiles (*Alligator sinensis*) include formants. To separate the acoustic effects of the source and the filter, they used helium to expand the respiratory system and thereby modify its resonance properties, which results in altered formant distributions while preserving the frequency bands of the source signal. This experiment provided the first evidence of formants in non-avian reptiles and suggests previously unexpected similarities in the vocalizations of species across the entire amniote clade. This clade includes the classes avian reptiles (hereafter called birds), non-avian reptiles (hereafter called reptiles), and mammals and is rooted in their last common ancestor, the stem-amniotes (see figure 3). The stem-amniotes diverged into the Synapsidia, today's mammals, and the Reptilia 331–319 mya (million years ago). The Reptilia subsequently diverged into the Captorhinidae and the Diapsida. The Neoreptilia later emerged from the Diapsida and in turn diverged into the Parareptilia and the Neodiapsida, the latter being the ancestor of the Archosaurs, turtles, and the Lepidosaurs (Ford & Benson, 2020). Archosaurs include today's crocodilians and birds, while Lepidosaurs include today's tuataras, snakes, and lizards (Zug, Vitt, & Caldwell, 2007). Based on a miRNA-based reconstruction, turtles are a sister group to Archosaurs and therefore belong to the Neodiapsida (Field et al., 2014). The focus here is on amniotes vocalizing on land, as the vocalizations of aquatic vertebrates under water face different physical constraints (see Ladich & Winkler, 2017) that would be outside the scope of this thesis.

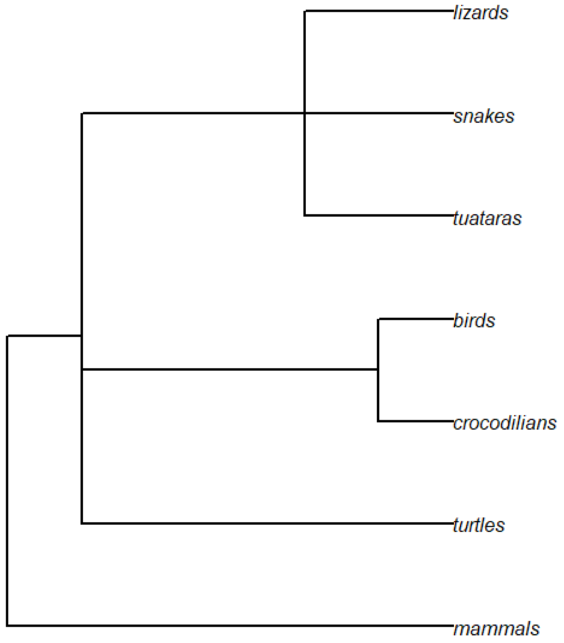


Figure 3: Simplified phylogenetic tree of amniotes pruned for the species discussed in this thesis. The tree was built in R (version 3.6.1, R Core Team, 2019) using the package ape (Paradis E. & Schliep K., 2018) version 5.3 and based on the data presented above. The outmost node to the left represents the last common ancestor of all mentioned species, a stem-amniote. Successive nodes represent the respective last common ancestor of the species within that section. Most recent species are on the far right.

1.2. Vocalization - Production

Reptiles use a larynx to produce sounds (Colafrancesco & Gridi-Papp, 2016). The larynx is a cartilaginous structure inside the pharyngeal cavity located at the junction of the trachea and the esophagus. It consists of two arytenoid and a cricoid cartilages that are supported by the hyoid (Sacchi, Galeotti, Fasola, & Gerzeli, 2004). The glottis is opened and closed by a dilator and a constrictor muscle. Most snakes only have a small larynx without vocal folds (Colafrancesco & Gridi-Papp, 2016), but there is at least one species that has vocal folds (Young, Sheft, & Yost, 1995). After the inhalation of helium, the growls of the mangrove ratsnake (*Gonyosoma oxycephalum*) show an increase in frequency. This frequency shift through expansion indicates that the trachea serves as a resonating chamber, creating a resonance effect (Young, 1991). Results of a similar experiment with heliox and crocodiles were interpreted as evidence of formants (Reber et al., 2015). Lizards possess elastic vocal cords with which they produce complex tone calls containing harmonic structures (X. Yu et al., 2011).

Birds possess both a larynx and a syrinx. The syrinx is used for sound production through vibrations of soft tissues comparable to the mammalian vocal folds. This vibratory tissue is located within the syringeal skeleton, either in a tracheal or bronchial location, around the tracheo-bronchial junction, or inside the bronchi (King 1989 as cited in (Düring & Elemans, 2016), (Riede & Goller, 2010a). The syrinx is a unique organ that likely evolved due to selective pressures on vocalizations as its sound production is more efficient than sound production from the laryngeal position (Riede, Thomson, Titze, & Goller, 2019). The labia consists of connective tissue (Riede & Goller, 2010a) and is controlled by syringeal muscles. Birds rely on the same myoelastic-aerodynamic mechanism to vocalize that humans use to speak, despite using the syrinx instead of the larynx for this purpose (Elemans et al., 2015). The upper vocal tract is made up of the trachea, (which is connected to the syrinx), the larynx (the end of the trachea), the oropharyngeal-esophageal cavity, and the beak. Tracheal length is correlated with the body size of the individual and so a tracheal elongation is used to exaggerate the vocalizer's size (Fitch, 1999). Despite being different organs, the larynx and the syrinx are nevertheless comparable, as they both exhibit myoelastic-aerodynamic characteristics and rely on the interplay of respiration, the vocal organ, and the vocal tract to produce sounds, as well as adduction, abduction, and layered, oscillating tissue (Riede & Goller, 2010b). Unlike the syrinx and the mammalian larynx, the avian larynx does not contain vibratory tissue (Häcke 1900, as cited in Düring & Elemans, 2016) and the structure of the laryngeal cartilages also differs between these taxonomic groups (Hogg, 1982). As in the reptilian vocal tract, the glottis is controlled by the dilator and the constrictor muscles (King 1993 as cited in Düring & Elemans, 2016). Details of the vocal tract structures may vary between species (Düring & Elemans, 2016). Zebra finches and large-billed crows thus produce their vocalizations with the syrinx instead of the larynx. Avian sound production systems are more suited for fast vocalizations, giving rise to the important temporal patterns found in birdsong (Riede & Goller, 2010b).

In mammals, the larynx consists of the thyroid cartilage, the cricoid cartilage, and two arytenoid cartilages. Vocal cords, or rather vocal folds, attached to the thyroid cartilage and the arytenoid cartilages produce the source signal. The expiration of air opens the glottis, the space between the two vocal folds (Titze, 1994). Humans can modify the shape of the sounds they produce by altering the resonance properties of the vocal tract through movement of the articulators, namely the tongue, velum, lips, and lower jaw (Fant 1960, as cited in Taylor et al., 2016), whereas other mammals usually have less movement control of their articulators and thus produce vocalizations that are more restricted and predictable in their shape (Fitch, 1994).

A small number of species can perform vocal mimicry and imitate the vocalizations of other species, including parrots, corvids, orangutans, and elephants. Corvids can learn to imitate human speech, such as Konrad Lorenz' pet raven Roah that could reportedly say his own name with a human intonation (Lorenz, 1952). A captive orangutan has been found to produce a novel vocalization that require vocal fold control exceeding species-typical levels (Lameira, Hardus, Mielke, Wich, & Shumaker, 2016). Asian elephants (*Elephas maximus*) can imitate human speech, match Korean formants and fundamental frequency to a degree that enables native Korean speakers to understand it (Stoeger et al., 2012).

On the molecular level, there are a considerable number of convergent genes involved in motor control and vocalization acquisition (Pfenning et al., 2014). The the transcription factor FOXP2 in humans, or FoxP2 in zebra finches plays a prominent role for song and speech acquisition. The expression patterns of FoxP1 and FoxP2 in zebra finches are also similar to those in human foetuses (Teramitsu, Kudo, London, Geschwind, & White, 2004). A mutation of FOXP2 can lead to severe language disorders in humans (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001). In zebra finches, FoxP2 is involved in song learning, and so a knockdown disrupts song imitation and results in incomplete and inaccurate songs, an abnormality that is comparable to those observed in humans (Haesler et al., 2007). Similarly, knockdown of FoxP1 and FoxP4 also results in impaired song learning in juvenile zebra finches (Norton, Barschke, Scharff, & Mendoza, 2019). FOXP2 has also been found in the genome of the American Crow (*Corvus brachyrhynchos*) (O'Leary et al., 2016).

1.3. Vocalization - Perception

It should be noted that voice recognition (individual identification) and voice discrimination (perception of a difference), despite sometimes being used synonymously, are distinct cognitive abilities and while the latter is a prerequisite for the former, the inverse is not the case. Support for this distinction has been obtained from studies on brain-damage in which impairment of one ability does not automatically include impairment of the other (van Lancker & Kreiman, 1987). Individual recognition broadly requires two steps: a signaller has to produce unique recognition cues, such as the learned song in zebra finches (Zann, 1996) or variations in formant frequencies based on physiological differences such as the voice cues in rhesus monkey calls (Rendall et al., 1998), and the receiver has to perceive these cues, learn and remember them in relation to the individual, and recall them in future interactions with the signaller (Tibbetts & Dale, 2007).

The principal hearing mechanism is shared across amniotes. When moved by incoming sound waves, the hair cells located in the cochlea in the inner ear are moved by the vibrations of the sound (Manley, 2000). If their movement reaches a certain threshold, they trigger the auditory nerve fibres to fire action potentials (electric signal, AP). The AP is then transmitted to the cochlear nucleus in the brainstem and the auditory cortex via the auditory nerve (Brownell & Manis, 2014; Fitch, 2018). In humans, the sound is captured by the outer ear (pinna), travels through the auditory canal and then sets the tympanic membrane in motion. This in turn moves the auditory ossicles (hammer, anvil, stirrup), the resulting vibrations are transferred through the oval window of the fluid-filled cochlea, where they will cause the inner ear cells in the scala media to move. Humans, like other mammals, also have outer hair cells that serve the amplification of sound (Manley, 2000; Yost, 2000).

Despite these structural similarities of sound perception both between amniote species and between individuals of the same species, perceptual abilities may vary. In humans, variation in phoneme categorization tasks has been found to correlate with differences in sensitivity to the relevant acoustic cues, f0 and voice onset time (Kong & Edwards, 2016). This variation may be due to differences in cue weighting (Holt, Tierney, Guerra, Laffere, & Dick, 2018), as well as encoding and mental representation (Ou & Law, 2017). Many insights into the genetic basis of acoustic perception come from research on disorders related to the processing of speech and other acoustic information (see for example Rice & Smolík, 2011), such as auditory processing disorder (APD) which inhibits patients from perceiving and processing sounds despite functional hearing (Bamiou, Musiek, & Luxon, 2001). Auditory processing skills affected by this disorder include frequency and temporal resolution of sounds, two skills that have been found to be heritable (Brewer et al., 2016). A disorder affecting auditory duration discrimination in several members of one family has been linked to a chromosome 12 locus (Addis et al., 2010). Some individuals are able to perceive absolute pitch (recognition without external reference pitch, Takeuchi & Hulse, 1991), an ability influenced both by musical background and genetics (Baharloo, Johnston, Service, Gitschier, & Freimer, 1998). This is not to say that speech disorders are exclusively heritable or that all aspects of sound perception are heritable, but rather aims to highlight that at least some abilities related to the perception of acoustic information are at least partially genetically determined and individual variation is to be expected.

Perceptual variation between species includes differences in perceivable frequency range (Manley, 2000) or the sensitivity to certain components within vocalizations. Songbirds are much more sensitive to details in temporal fine structure cues than humans, indicating that how humans and songbirds hear birdsong may differ greatly (Dooling & Prior, 2017). Instead of the sequence of song syllables, they are more sensitive to the acoustic features contained within each individual syllable (Fishbein, Idsardi, Ball, & Dooling, 2020). As formant dispersion is less accurate in high-frequency vocalizations (Fitch, 1997), signature songs in zebra finches may be an adaption to this constraint (Elie & Theunissen, 2018). Nevertheless, they are still sensitive to formant patterns in human speech (Ohms, Gill, van Heijningen, Beckers, & ten Cate, 2010) and use the same cue-weighting bias as Dutch adults as they prefer higher formants over lower formants to discriminate vowels (Ohms, Escudero, Lammers, & ten Cate, 2012), use vocal tract filtering to modulate their vocalizations (Ohms, Snelderwaard, ten Cate, & Beckers, 2010) and show evidence of perceptual constancy necessary to normalize variation in song

(Pike & Kriengwatana, 2019). However, zebra finches are not the only animals who perceive certain elements of human language: Chinchillas discriminate between the voiced and voiceless plosive consonants [t] and [d] (Kuhl & Miller, 1975), cotton-top tamarin monkeys (*Saguinus Oedipus*) and long-evans rats (*Rattus norvegicus domestica*) can discriminate Dutch and Japanese sentences by their prosodic features (Ramus, Hauser, Miller, Morris, & Mehler, 2000; Toro, Trobalon, & Sebastián-Gallés, 2003), and large-billed crows can discriminate between a familiar and unfamiliar language without prior training (Schalz & Izawa, 2020).

In turn, humans exhibit sensitivity to vocalizations of multiple heterospecific taxa as well. Speakers of multiple language groups perceive emotional arousal in the vocalizations of amphibians, avian and non-avian reptiles, and mammals (Filippi et al., 2017). Infants may equally profit from non-human primate vocalizations as they do from human speech when it comes to object categorization. Ferry, Hespos and Waxman (2013) found that both lemur vocalizations and human speech support object categorization in three and four-months-old infants, while object categorization in six-months-olds is only promoted by human speech. Backward speech has not been found to be beneficial for either age group. The initially present advantage of non-human primate vocalizations can be maintained through exposure to them, which suggests that both types of vocalizations but not backward speech rely on an “initial template” that is then modified through experience (Perszyk & Waxman, 2016). One reason for this may be that backward speech is not typically associated with communication. Sine wave tones can support object categorization, if they had previously been presented in a communication setting (Ferguson & Waxman, 2016). This suggests that infants perceive both human speech and non-human primate vocalizations as communicative, which points towards underlying parallels that are not present in backward speech or sine wave tones. Infants’ initially equal preference for human speech and non-human primate voices provide further indication for underlying parallels between these vocalizations. Right after birth, neonates equally prefer human speech and the vocalizations of rhesus monkeys (*Macaca mulatta*) over synthetic sounds, but will prefer human speech over both rhesus vocalizations and synthetic sounds three months later (Vouloumanos, Hauser, Werker, & Martin, 2010). However, this preference seems to be quite flexible, as another study found that 9-months-old infants prefer monkey vocalizations over human speech, while still preferring speech over synthetic non-speech sounds (Sorcinelli, Ference, Curtin, & Vouloumanos, 2019).

Friendly, Rendall and Trainor (2014) showed that six-months-old infants can discriminate two rhesus monkeys by their voices. They compared adults (between 18 and 40 years of age) with infants aged 5.5 to 6.5 months, and infants aged 11.5 to 12.5 months in their ability to either discriminate between the voices of two female Canadian speakers, or the “coo” calls produced by two rhesus monkeys. Adults were tested in a Same-Different-Task, while infants were given a Conditioned Head Turn Task. Adult participants' performance with human voices was far better than with rhesus voices, although they were still able to discriminate both at above chance levels. Adults also performed better with human voices than infants, while the younger infant group outperformed both the older infant group and the adult group in discriminating the rhesus monkey voices. 12-months-old infants already showed increased difficulty with the discrimination of rhesus monkeys, while their sensitivity to differences between human voices had increased compared to the younger infant group. The authors conclude that the sensitivity to individual differences of heterospecific voices is subject to perceptual narrowing, and that while humans are equally perceptive to human and non-human primate at birth, their unilateral experience with conspecifics leads to an increased sensitivity to human voices, and a decreased sensitivity to those of other, less relevant species. However, the ability to discriminate rhesus monkeys' voices remains somewhat plastic even after perceptual narrowing, and with practise 12-months-old infants can re-gain this ability and perform as well as six-months-olds (Friendly, Rendall, & Trainor, 2013). To my knowledge, no voice discrimination experiment with human participants has been conducted with vocalizations of any non-primate species yet, but it has been shown that infants as young as five months can discriminate between the innate, lower frequency calls of sea birds and the learned, high-frequency songs of garden birds. While the task was completed successfully by all age groups, infants aged five to seven months showed a preference for the sea bird calls, infants aged 10-12 months displayed a decreased preference for the sea bird calls, and adults preferred the garden bird songs over the sea bird calls altogether (Lange-Küttner, 2010).

The evidence presented here suggests that "[i]t might not be the case that the genetic code 'generated a vast number of mutually incomprehensible communication systems across species while maintaining clarity of comprehension only within a given species' (Hauser et al., 2002) and that infants' "[i]nitial biases, and rapid attunement, may constitute conserved and fundamental principles that underwrite the development of perceptual systems across the animal kingdom" (Vouloumanos et al., 2010).

The following series of experiments therefore aims to examine whether humans can discriminate individual zebra finches and individual large-billed crows by their respective individual vocalizations. Particular attention is given to potential differences in discrimination accuracy based on participants' sex and musical background, as well as potential improvement over time and potential differences in discrimination accuracy between the vocalizations of these two species. Discrimination is expected to be possible with both species, however discrimination accuracy is expected to be higher with zebra finch songs as opposed to crow calls. While large-billed crows have a "signature voice system" that also contains pitch contour patterns with relatively little intra-individual variation (Kondo, Izawa, & Watanabe, 2010) their calls are not as stereotyped as the learned signature songs produced by zebra finches (see Appendix A, figures 3 and 5). The relative importance of this cue will be directly assessed in an additional experiment using zebra finch song without pitch contour. Men are expected to perform better than women, as men have been found to more accurately estimate an animal's body size based on formant dispersion (Charlton et al., 2013). While size estimation and individual discrimination are different tasks, a higher sensitivity for formant dispersion may also provide an advantage in the present experiment. Additionally, participants with a musical background are expected to gain higher scores than non-musical participants, because musicians are more sensitive to global voice timbre (Chartrand & Belin, 2006) and pitch (Pitt, 1994). The increased sensitivity to differences in these cues might be transferable to this task and thus provide an advantage. Their potential influence on performance differences between participants in this study may provide some insights into what acoustic cues and perceptual abilities are involved in this task. As the main purpose of this experiment is to establish whether this discrimination is possible at all, participants were not screened for their musical background. As such, the number of participants who report a musical background is too low to make definite conclusions about the relevance of this characteristic and the results reported here only aim to provide moderate suggestions with the available data. A formal investigation of its relevance would be required in future experiments to make a definite statement. A performance trend then indicates whether the discrimination is more likely based on a learnt skill or an innate trait. It should be noted that I'm not implying that this "skill" or "trait" is exclusively used for the discrimination of zebra finches or crows, but rather a general perceptual ability that can additionally be used for this purpose, amongst other functions. I do not think that there is a "zebra finch discrimination" – ability, but instead one or multiple perceptual mechanisms that respond to to one or multiple features used for this discrimination, as well as in other domains (especially language and music).

2. Experiment 1: Discrimination of Individual Zebra Finches

2.1. Material and Methods

Subjects

An adult sample was used (N=50, 25 female). Sample size was based on an a-priori power analysis using GPower version 3.1 (Erdfelder, Faul, & Buchner, 1996) with an estimated effect size d of 0.85 and a group allocation ratio of 1 for a minimum power of 0.8 in a two-tailed Mann-Whitney U test (used to analyse the correlation between participants' sex and discrimination accuracy, see analysis), while being sufficiently large (power=0.8) for a two-tailed, one-sample Wilcoxon signed-rank test to test whether discrimination accuracy was above chance level. The effect size was cautiously estimated with the calculated effect size of 1.27 reported by Charlton et al. (2013) in their analysis of a potential correlation between sex and acoustic size judgement. Subjects were students and staff at Middlesex University between the ages 18 to 50 (mean age unknown). Age-related loss of sensitivity to frequency differences should not be an issue with participants younger than 65 (Harris, Mills, He, & Dubno, 2008). Participants did not report hearing problems and gave informed consent. Approval of the ethics application had been obtained prior to data collection (see Appendix B, figure 1). 16 participants reported to have a musical background. No participants were removed before the analysis.

Stimuli

Stimuli consisted of song elements of two different zebra finches (both male, 3 and 4 months old) that I recorded at Bielefeld University (shotgun microphone, sampling rate of 44100Hz). Animal housing and song recording were in compliance with all applicable national guidelines for the care and use of animals, and recording was authorized by Prof. Barbara Caspers. Recordings were analysed in Praat version 6.0.49 (Boersma & Weenink, 2019).

The domestication of the Australian zebra finch during the last 200 years (Immelmann 1965, as cited in Slater & Clayton, 1991) has led to behavioural differences between the wild type and captive individuals. Compared to wild individuals, zebra finches from the Bielefeld colonies were found to produce longer phrases (0.14s longer on average), show more element variety (1.8 elements more on average) and sing faster songs (0.9 elements per second more on average), possibly because breeders selected for more complex calls (Slater & Clayton, 1991). This data was collected almost 30 years before the recordings for the present experiments were made, so further divergence from the wild type is to be expected, although the extent of that divergence is unclear and a recent study found no evidence of divergence in genetically determined constraints (Lachlan et al., 2016). Differences are therefore likely superficial but should be kept in mind in the following.

Nine motifs per zebra finch were selected based on maximum similarity to the bird's individual stereotypical pattern indicated by pitch contour, intensity contour, duration and arrangement of elements (see table 1 for mean values, and Appendix A, table 1 for values of each stimulus). Each motif was high-pass filtered at 500Hz with the software Audacity version 2.3.0 (Audacity Team, 2019) to reduce low-frequency background noise of the bird clanging the perch against the metal cage bars without interfering with the high-frequency song. Zebra finch songs naturally vary in the number, length, frequency and amplitude patterns of motif elements and so these features were not standardised to keep the stimuli as naturalistic as possible. Introductory elements were not included, as they do not contain discrimination cues (Sossinka & Böhner, 1980; Zann, 1996). A silent 2s interval was added at the end of each motif to create clear breaks between them (see table 1 for mean acoustic features and figure 4 and 5 for spectrograms of each zebra finch song). As Praat could not extract reliable values for the third formant F3 was not further analysed for the two zebra finch experiments (see high standard variation for F3 in table 1, and Appendix A, figures 1 and 2).

Table 1: Acoustic features of the nine selected motifs of each zebra finch. Frequency range was set to a minimum 50Hz and maximum 10,000Hz for the pitch analysis (note that Praat measures pitch instead of F0) and to a maximum 10,000Hz and 3 extracted formants (see figures 4 and 5 for the spectrograms, and Appendix A, figures 1 and 2 for the formant analysis).

	Zebra finch *A* mean	**Zebra finch *A* SD**	**Zebra finch *B* *mean***	**Zebra finch *B* SD**
Duration per motif (ms)	397.6	10	335.5	8
Intensity per motif (dB)	59.7	0.5	59.1	1.5
Pitch per motif (Hz)	3177.4	204	2888.8	309.3
Frequency of the first formant (Hz)	3183.4	33.2	3368.5	73.5
Frequency of the second formant (Hz)	4743.6	53.6	5177.4	91.9
Frequency of the third formant (Hz)	6884.1	44.1	7268.8	341.5
F1-F2 dispersion (Hz)	1560.6	38.4	1807	54.7

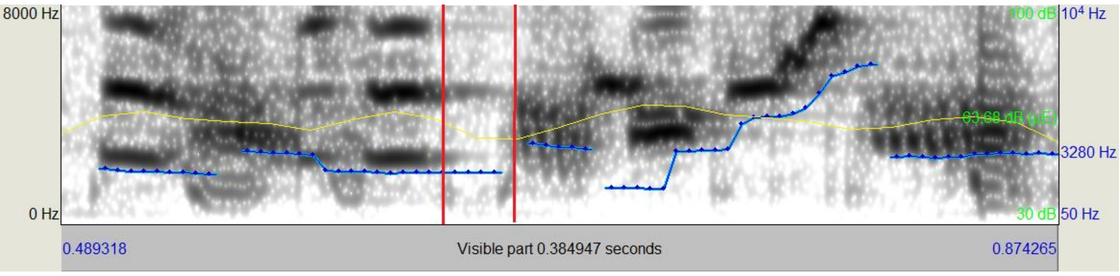


Figure 4: Spectrogram of one song motif of zebra finch *A*, created in Praat version 6.0.49 (Boersma & Weenink, 2019). Red lines indicate element boarders, blue lines indicate pitch (pitch values written in blue on the right side in Hz), and yellow lines indicate intensity (intensity values written in green on the right side in dB). Song motifs consisted of two alternating elements.

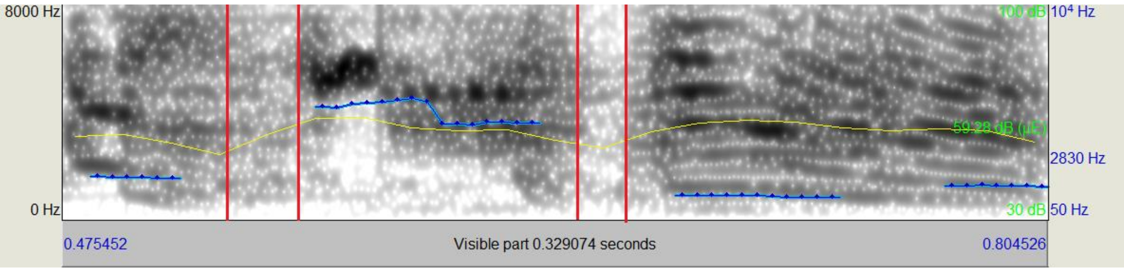


Figure 5: Spectrogram of one song motif of zebra finch *B*, created in Praat version 6.0.49 (Boersma & Weenink, 2019). Red lines indicate element boarders, blue lines indicate pitch (pitch values written in blue on the right side in Hz), and yellow lines indicate intensity (intensity values written in green on the right side in dB). Song motifs consisted of a sequence of three different elements.

Apparatus

The participant background questionnaire and discrimination task was presented in the software PsychoPy version 3.2 (Peirce et al., 2019) on a desktop computer in a quiet room. Sounds were played over headphones.

Procedure

Participants were tested with the forced-choice Same-Different Paradigm (Pisoni & Lazarus, 1974) consisting of 40 trials following the previous study on the discrimination of rhesus monkey voices by humans (Friendly et al., 2014). Each trial contained two vocalizations, either produced by the same individual ("same"-trial) or two different individuals ("different"-trial). The stimuli and their combination ("same" or "different") were chosen at random each trial to avoid predictability. Participants were not told what species they would listen to. Before the first trial, they received the following instructions: "You will now hear 40 sound pairs. A pair of sounds was either produced by the same animal or by two animals of the same species. After each pair, you will be asked to decide whether you heard the same animal or two different animals. Sounds are separated by a 2s interval and only 0.3s long." After the playback of each pair, participants were asked “Was the song sung by the same bird?”, to which they could reply either yes (keypress “y”) or no (keypress “n”). During the experiment, participants did not receive feedback on the correctness of their decisions.

Analysis

Key responses were recorded in PsychoPy (Peirce et al., 2019) and were divided into four response categories: Hit (y on a “same”-trial), miss (n on a “same”-trial), correct reject (n on a “different”-trial), and false alarm (y on a “different”-trial; see table 2).

Table 2: Response category matrix for the participants’ possible key responses (“yes” and “no”) in relation to the respective trial (“same-trial” and “different-trial”).

	Yes	**No**
Same-trial	Hit	Miss
Different-trial	False Alarm	Correct Reject

The hit rate (proportion of hit responses out of all same-trials) and the false alarm rate (proportion of false alarm responses out of all different-trials) were used to calculate the discrimination sensitivity index d’ formalized in the signal detection theory (Stanislaw & Todorov, 1999) in R (R Core Team, 2019) using the R package psyphy and the command “dprime.SD(H, FA, method = "diff")” (Knoblauch, 2014). A d’ score of 0 indicates that participants were unable to discriminate the two stimuli sets, higher values indicate more accurate discrimination performances. Note that there is no categorization into successful and unsuccessful discrimination, but rather a continuous scale of more or less accurate discrimination. As d’ scores cannot be negative, a “false alarm” rate that was higher than the corresponding “hit” rate was given a d’ score of 0. As d’ scores cannot be calculated with perfect “hit” and “false alarm” rates of 1 and 0, these rates were corrected with formula 1 as described by (Snodgrass & Corwin, 1988). This was done seven times for rates of 1 and 12 times for rates of 0. Three single trials were missing and thus not included in the analysis.

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

Mean d’ scores were calculated for each condition. Since the d’ score data was not normally distributed (assessed with a Shapiro test in R), non-parametric tests were chosen for this part of the analysis. Whether d’ scores were significantly above chance level was determined with a one-sample Wilcoxon signed-rank test. D’ scores were also compared between male and female participants, and participants with and without a musical background using a Mann-Whitney-U test.

To assess which acoustic features were most important, the difference in mean pitch frequency, mean frequency of the first formant (F1), and frequency dispersion between F1 and F2 were compared between stimuli pairs that triggered a "false alarm" response and those that triggered a "miss" response. This is based on the assumption that pairs similar in frequency should trigger "false alarm" responses more often, while pairs with high differences should prompt the response "miss" more frequently, if these cues are in fact relevant. These acoustic features were chosen because pitch is the most important cue to determine the difference between two human speakers, followed by mean F1 and formant dispersion (Baumann & Belin, 2010) and participants may rely on this ability for non-human vocal discriminations as well. F3 was not considered due to this cue's high variation in zebra finch *B* (SD=341.5, see table 1). Due to the large volume of stimuli pairs (320 mistakes in total) for which individual pitch differences, F1 differences, and dispersion differences would have to be calculated separately, only pairs that triggered the same response most often were considered in this analysis. This also minimizes the confounding effect of pairs that are not in fact difficult to discriminate and only triggered a mistake once or twice due to other reasons (such as lack of attention). The threshold for this analysis was therefore chosen to include pairs with the highest frequency of triggered mistakes while also including enough pairs for the analysis. Based on the frequencies of pairs in this experiment, the threshold of minimum frequency of triggered mistake types was three, regardless of stimuli order within the pair (e.g. the pairs a2b3 and b3a2 were considered to be the same). A total of 23 "false alarm" pairs and 22 "miss" pairs were selected (see Appendix A, table 2). Every pair was counted only once in the analysis regardless of how often it triggered a mistake (since they were already selected based on frequent appearance). Participants who scored a d' value of 0 were not included as they did not perceive any difference between the stimuli (1 participant excluded).

A potential learning trend throughout the experiment was analysed with a linear regression model (lm(percentage correct ~ trial number)) to see whether the discrimination success improves over the course of the experiment. The average number of trials needed before reaching the first success streak (threshold of 5 correct answers in a row, even if followed by further mistakes, not including 1 participant with d'=0. Two participants who had no success streak were counted as needing all 40 trials) and the average number of mistakes across all trials were calculated as additional insights into possible individual learning patterns.

The confidence interval was calculated in R based on mean (a), standard deviation (s) and sample size (n) using the formula qt(0.975,df=n-1)*s/sqrt(n) for the error and a+-error for the interval margin. The effect size Cohen's d was calculated with formula 2. It indicates the difference between two values (in this case the expected d’ score “mu” and the sampled d’ score) where d=0.2 is considered a small effect size and d=0.8 a large effect (J. Cohen, 1988). Power was calculated with GPower (Erdfelder et al., 1996) version 3.1.

$$2)\ d = \frac{mean\ (data) - mean\ (mu)}{standard\ deviation}$$

2.2. Results

The average d’ score was 3.68 (SD=1.54, 95% CI [3.24, 4.11]) and individual scores ranged from 0 to 5.94, the highest possible score. D’ scores were significantly above chance level (one-sample Wilcoxon signed-rank test with an expected d’ score of 0, V=1225, d=2.38, power=1, p<0.01).

A Mann-Whitney U test showed no significant difference in d’ scores between female and male participants (W=268; see figure 6).

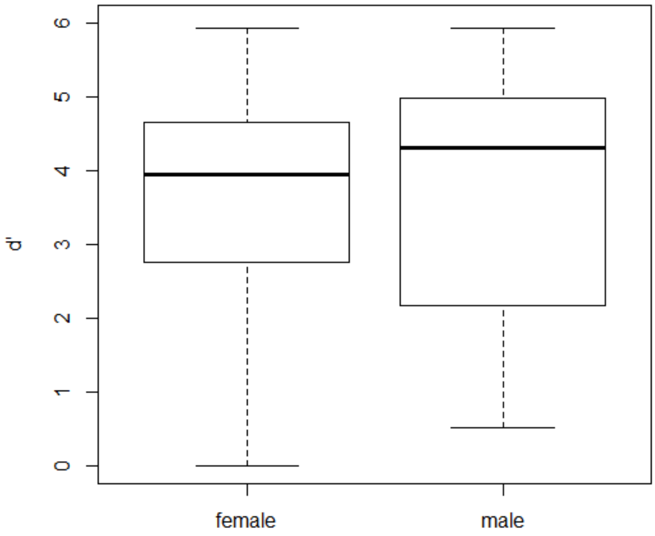


Figure 6: d’ scores obtained by 25 female and 25 male participants, created in R version 3.6.1 (R Core Team, 2019). Solid lines indicate median d’ scores, the box contains the 25% of values above and below the median, and the whiskers the outmost 25% of values above and below the inner 50% box. Outliers (single sample points outside the whiskers) would be indicated as circles but are not found in this sample. Mean d’ scores are 3.54 and 3.83 respectively.

A Mann-Whitney-U test showed no significant difference between participants with and without musical background (W=180.5; see figure 7).

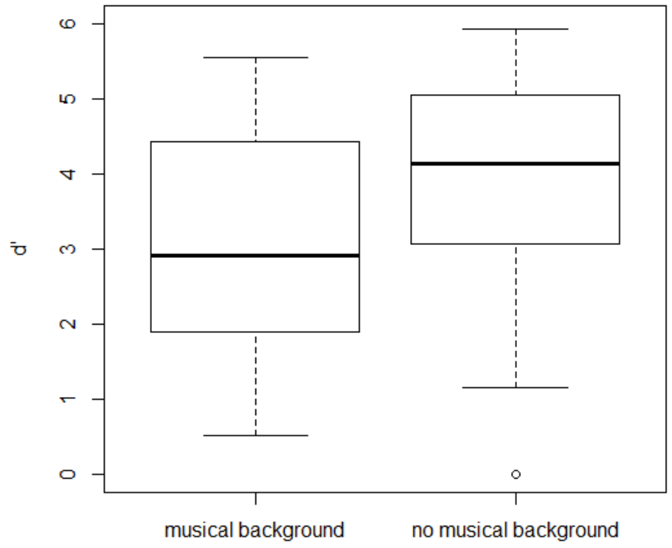


Figure 7: d' scores obtained by 16 participants with a music background and 34 participants without a musical background, created in R version 3.6.1 (R Core Team, 2019). Mean d' scores are 3.09 and 3.96 respectively.

14 "miss" responses out of 157 were triggered by pairs with identical stimuli (e.g. a6a6). A Mann-Whitney-U test showed no significant difference between pitch differences that triggered "false alarm" responses and those that triggered "miss" responses (W=0.22), but did show a significant difference for both F1 differences (W=450, p<0.01) and formant dispersion differences (W=506, p<0.01; see figure 8).

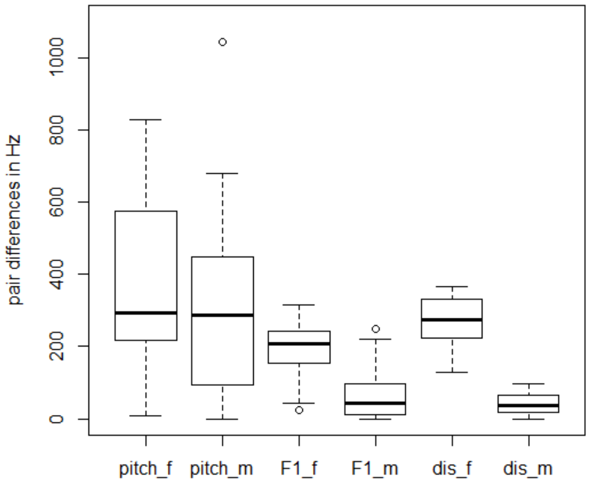
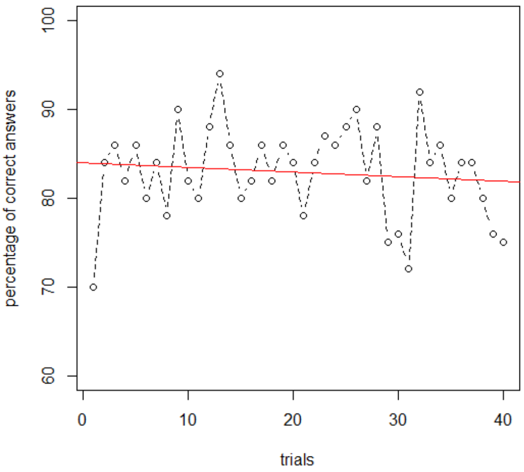


Figure 8: pitch, F1, and dispersion between F1 and F2 differences within pairs that triggered a "false alarm" response ("_f") or a "miss" response ("_m"), created in R version 3.6.1 (R Core Team, 2019). Mean frequencies were 382Hz, 309Hz, 192Hz, 62Hz, 270Hz, and 40Hz respectively.

The average participant needed 5 trials (SD=8.9) before reaching the first success streak (threshold of 5 correct answers in a row) and made 7 mistakes (SD=5.4) in total. A linear regression analysis showed no significant trend across all trials (m=-0.04; see figure 9), as well as no significant, although positive trend across the first 20 trials (m=0.25; see figure 10). The sub-group of participants with a below-average d' score shows a steeper decline in the percentage of correct answers per trial compared to the entire sample group (m=-0.09).



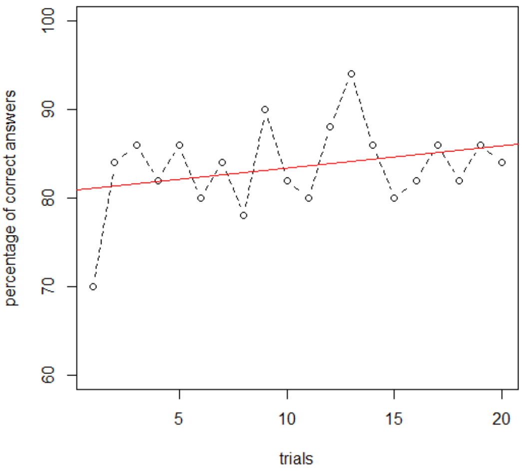


Figure 9 (left): Percentage of correct answers (either "hit" or "correct reject") for each trial with one answer per participant and trial and a linear regression line indicating overall trend, created in R version 3.6.1 (R Core Team, 2019).

Figure 10 (right): Percentage of correct answers (either "hit" or "correct reject") for each of the first 20 trials with one answer per participant and trial and a linear regression line indicating overall trend, created in R version 3.6.1 (R Core Team, 2019).

3. Experiment 2: Discrimination of Individual Zebra Finches Without Pitch Contour

3.1. Material and Methods

Subjects

Adults (N=25, 14 female). The sample size was based on an a-priori power analysis using GPower version 3.1 (Erdfelder et al., 1996) with an allocation ratio of 2, an estimated effect size d of 0.85 for a minimum power of 0.85 with a two-tailed Mann-Whitney U test (used to analyse discrimination accuracy difference between experiments 1 and 2, see analysis), while being sufficiently (power=0.8) large for a two-tailed, one-sample Wilcoxon signed-rank test to test whether discrimination accuracy was above chance level. The lower sample size reflects that this experiment is an extension to experiment 1 instead of an independent experiment and only serves to test the expected relevance of pitch contour on the discrimination accuracy. Subjects were students at Middlesex University between the ages 18 to 40 (mean age unknown), did not report hearing problems and gave informed consent. Approval of the ethics application had been obtained prior to data collection (see Appendix B, figure 1). Seven participants reported a musical background. No participants were removed before the analysis.

Stimuli

Stimuli from experiment 1 were manipulated and analysed in Praat version 6.0.49 (Boersma & Weenink, 2019). Pitch contour was eliminated by removing existing pitch points and adding new pitch points at the mean frequency of the corresponding motif to keep the differences in overall frequency (see table 2 for mean acoustic features, and Appendix A, table 3 for acoustic features per stimulus) at the time points 0.0001s, 0.1s, 0.2s, 0.3s, and 0.4s. Each sound file was then checked, and additional pitch points were added to achieve a flat pitch contour line (see Appendix A, figure 4).

Table 2: Acoustic features of the nine manipulated motifs of each zebra finch. Frequency range was set to a minimum 50Hz and a maximum 10,000Hz for the pitch analysis and to a maximum 10,000Hz and 3 extracted formants (indicated by the spectrograms) for the formant analysis.

	Zebra finch *A* mean	**Zebra finch *A* SD**	**Zebra finch *B* mean**	**Zebra finch *B* SD**
Duration per motif (ms)	397.6	10	335.5	8
Intensity per motif (dB)	59.7	0.5	59.1	1.5
Pitch per motif (Hz)	3124	192.6	2872.5	304.7
Frequency of the first formant (Hz)	3190.2	171.4	2962.7	222
Frequency of the second formant (Hz)	5517.7	223.5	5563.1	268.6
Frequency of the third formant (Hz)	7188.4	211.3	7676.6	306.9

Apparatus

The set-up was identical to experiment 1.

Procedure

The experimental procedure was identical to experiment 1.

Analysis

Calculation and analysis of d’ scores were identical to experiment 1. One “hit” score was corrected. Due to the lower sample size, musical background was not analysed.

To examine which acoustic features might play a role in the discrimination, stimuli pairs were ordered according to the answer type they received (hit, miss, correct reject, false alarm). I chose to include all answer types instead of analysing only “false alarm” trials as in experiment 1 in consideration of the lower sample size and lower number of correct trials. In experiment 1, the majority of answers were correct and analysing correct trials would not have provided much information, whereas in experiment 2 a correlation between categorization success and certain pairs is more likely observable. The frequency threshold of pairs was raised to 4 for incorrect responses (14 “false alarm” and 16 “miss”) and to 5 for correct responses (18 “hit” and 17 “miss”) to provide a sample of the most frequent pairs similar in size to the sample in experiment 1. Formant dispersion for F1 and F2 was highly irregular (see Appendix A, table 3) and therefore not considered a potential cue. A Levene-test for variance homogeneity was conducted with the R package “car” (Fox & Weisberg, 2019), followed by a One-Way ANOVA. Four participants who scored a d’ score of 0 were not included.

A potential learning trend throughout the experiment was analysed with a Mann-Kendall trend test (as percentages of correct responses were not normally distributed) to see whether the discrimination success improves over the course of the experiment. The Mann-Kendall trend test was executed with the R package “Kendall” (McLeod, 2011). For the average number of trials needed to reach five correct answers in a row, six participants who had no success streak were counted as needing all 40 trials. Four were excluded due to a d’ score of 0.

The confidence interval was calculated in R based on mean (a), standard deviation (s) and sample size (n) using the formula qt(0.975,df=n-1)*s/sqrt(n) for the error and a+-error for the interval margin. Cohen’s d for the one-sample Wilcoxon signed-rank test was calculated with formula 2 (see experiment 1). Cohen’s for the Mann-Whitney U test was calculated with formula 3 were data 1 and data 2 refer to the different experiments, and the pooled standard deviation was calculated with formula 4 (J. Cohen, 1988). Power was calculated with GPower (Erdfelder et al., 1996) version 3.1.

$$3)\ d = \frac{mean\ (data\ 1) - mean\ (data\ 2)}{pooled\ standard\ deviation}$$

$$4)\ SDpooled = \sqrt{\frac{SD1^2 + SD2^2}{2}}$$

3.2. Results

The average d’ score was 1.3 (SD=0.82, 95% CI [0.96, 1.63]) and individual scores ranged from 0 to 3.29. D’ scores were significantly above chance level (one-sample Wilcoxon signed-rank test with an expected d’ of 0, V=231, d=1.58, power=1, p<0.01). A Mann-Whitney-U-test also showed a significant difference in d’ scores between experiment 1 (pitch contour) and experiment 2 (no pitch contour; W=131.5, p<0.01; d=1.56, power=0.99, see figure 11), but no significant difference between female and male participants (mean d’ scores were 1.38 and 1.19 respectively).

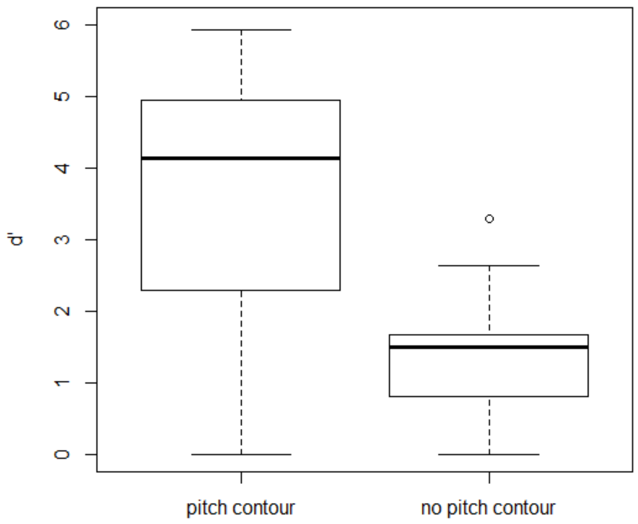
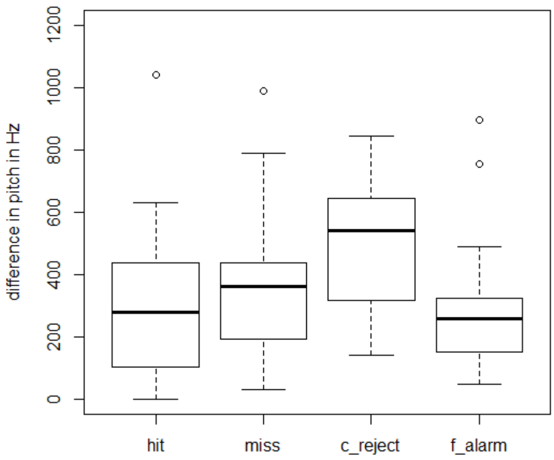


Figure 11: D' scores obtained by 50 participants in experiment 1 (with pitch contour) and 25 participants in experiment 2 (without pitch contour), created in R version 3.6.1 (R Core Team, 2019). Mean d' scores were 3.68 and 1.3 respectively.

Three stimuli pair triggered both "hit" and "miss" multiple times, and one pair triggered both "correct reject" and "false alarm" multiple times (see Appendix A, table 4). A One-Way ANOVA found no significant difference in either pitch frequency difference nor in F1 frequency difference between response types (F=2.33, see figure 12; F=2.6, see figure 13).



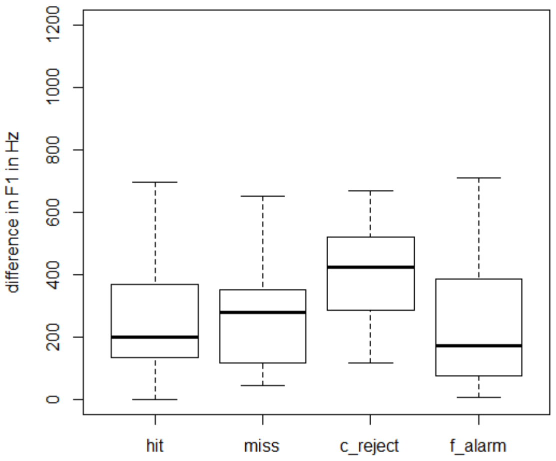
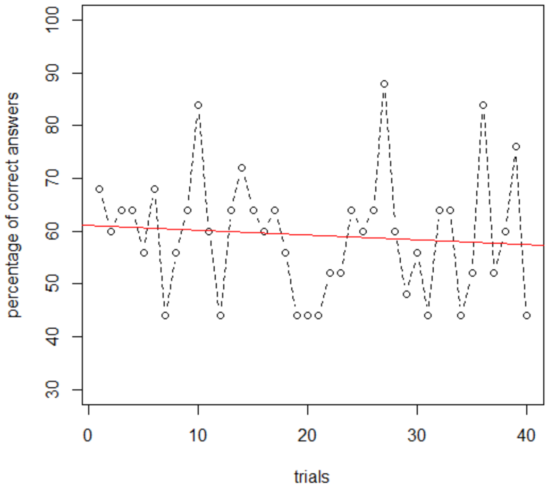


Figure 12: Difference in pitch frequency of each stimulus pair in Hz that triggered the same response (hit, miss, correct reject, false alarm) most often, created in R version 3.6.1 (R Core Team, 2019). Mean pitch differences were 307Hz, 368Hz, 500Hz, and 308Hz respectively.

Figure 13: Difference in F1 frequency of each stimulus pair in Hz that triggered the same response (hit, miss, correct reject, false alarm) most often, created in R version 3.6.1 (R Core Team, 2019). Mean F1 differences were 248Hz, 271Hz, 401Hz, and 246Hz respectively.

The average participant needed 21 trials (SD=15) before reaching the first success streak (threshold of 5 correct answers in a row) and made 16 mistakes (SD=2.9) in total. A Mann-Kendall trend test showed no significant trend across all trials (τ=-0.12; score=-94; see figure 14) and no significant trend across the first 20 trials (τ=-0.23; score=-42; see figure 15).



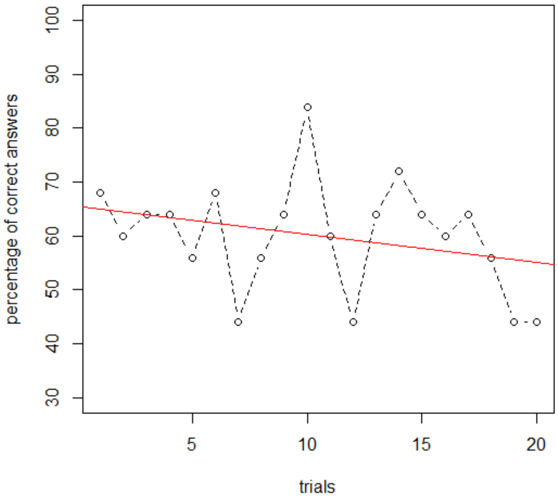


Figure 14: Percentage of correct answers (either “hit” or “correct reject”) for each trial with one answer per participant and trial, created in R version 3.6.1 (R Core Team, 2019).

Figure 15: Percentage of correct answers (either “hit” or “correct reject”) for each of the first 20 trials with one answer per participant and trial, created in R version 3.6.1 (R Core Team, 2019).

4. Experiment 3: Discrimination of Individual Large-Billed Crows

4.1. Material and Methods

Subjects

Adults (N=50, 32 female, 17 male). Sample size was based on an a-priori power analysis using GPower version 3.1 (Erdfelder et al., 1996) with an initially estimated effect size d of 0.85 for a minimum power of 0.8 with a two-tailed Mann-Whitney U test (used to analyse the potential correlation between participants' sex and discrimination accuracy in the previous experiments. The parametric counterpart, a two-tailed, unpaired t-test, works with a smaller sample), while being sufficiently (power=0.8) large for a two-tailed, one-sample Wilcoxon signed-rank test (or one-sample t-test) to test whether discrimination accuracy was above chance level. The group allocation ratio of the initial analysis was 1 but could not be realized due to lack of male participants. An updated a-priori power analysis with the actual allocation ratio of 1.88 and a total sample size of 50 showed that a minimum power of 0.8 for an effect size of 0.85 could be maintained since an unpaired t-test was used. Subjects were students and staff at Middlesex University between the ages 18 to 50 (mean age unknown), did not report hearing problems and gave informed consent. Approval of the ethics application had been obtained prior to data collection (see Appendix B, figure 1), as well as approval of the ethics amendments application (see Appendix B, figure 2). First-year Psychology students received credit points for participation (N=29). 11 participants reported to have a musical background. No participants were removed before the analysis.

Stimuli

Stimuli consisted of single ka-calls of two different Japanese large-billed crows (*Corvus macrorhynchos japonensis*, both female and 4 years old) that I recorded at Keio University, Tokyo (Sony HDR CX535 with build-in microphone, sampling rate 44100). The crows were alone in an outside aviary, and the ka-calls were likely of a territorial nature directed at nearby wild crows. Nine calls per crow were selected based on quality and lack of background noise from the recordings using Audacity version 2.3 (Audacity Team, 2019) and analysed in Praat version 6.0.49 (Boersma & Weenink, 2019). A silent 2s interval was added at the end of each call to create clear breaks between them (see table 3 for mean values and Appendix A, table 5 for details of acoustic features). Ka-calls of this species usually have a fundamental frequency of approximately 400Hz and are 0.3-0.4s long (Kondo, Watanabe, & Izawa, 2010), but as crow *A* only produced very short calls, the calls produced by crow *B* had to be shortened for them to match in duration (see table 3 for mean acoustic features and figures 16 and 17 for spectrograms of one of each crow's call). Kondo, Izawa, and Watanabe (2010) proposed that large-billed crows have a “signature voice system” for the purpose of vocal individual recognition based on low intra-individual and high inter-individual variation of acoustic cues that I will summarize in the following as mean pitch (see table 3) and pitch contour (see Appendix A, figure 5).

Table 3: Acoustic features of the nine selected motifs of each crow. Frequency range was set at 50Hz to 1,000Hz for pitch and at a maximum of 5,000Hz for formants.

	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
Frequency of the third formant (Hz)	4542.3	87	4148	124.9

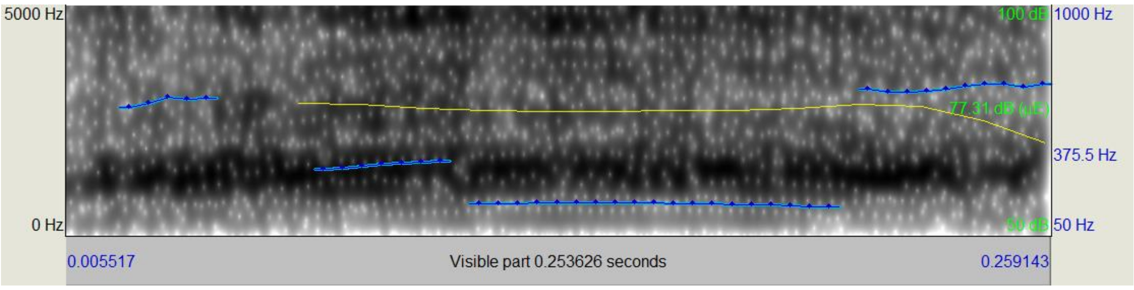


Figure 16: Spectrogram of one ka-call of crow *A*, created in Praat version 6.0.49 (Boersma & Weenink, 2019). Blue lines indicate pitch (pitch values written in blue on the right side in Hz), and yellow lines indicate intensity (intensity values written in green on the right side in dB). The mean pitch increase between the low pitch line and the high pitch line on the far right is 355.2 Hz (SD=92.6; see Appendix A, table 6).

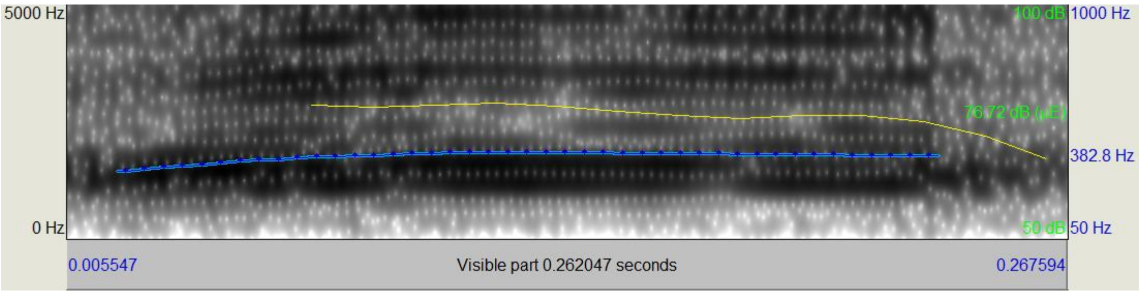


Figure 17: Spectrogram of one ka-call of crow *B*, created in Praat version 6.0.49 (Boersma & Weenink, 2019). Blue lines indicate pitch (pitch values written in blue on the right side in Hz), and yellow lines indicate intensity (intensity values written in green on the right side in dB).

Apparatus

The set-up was identical to experiment 1 and 2.

Procedure

The experimental procedure was identical to experiment 1 and 2.

Analysis

Calculation of d’ scores was identical to experiment 1. Three “hit” scores and five “false alarm” scores were corrected. Since d’ scores were normally distributed, parametric tests were used. D’ scores obtained here were compared with those from the non-normally distributed experiments 1 and 2 with a Mann-Whitney U test. Whether d’ scores were significantly above chance level was determined with a one-sample t-test. A Levene-test for variance homogeneity was conducted with the R package “car” (Fox & Weisberg, 2019) and showed no variance homogeneity for either the variable sex, nor the variable musical background. As such, both were analysed using an unpaired Welch t-test.

The analysis of stimuli pairs in relation to pitch frequency and F1 frequency, and F1-F2 dispersion was identical to experiment 1. Two participants were excluded from this analysis because their d’ scores were 0. The frequency threshold for pairs was five, resulting in a total of 17 “false alarm” and 21 “miss” pairs. An additional analysis was conducted regarding the relevance of pitch contour. Based on the pitch contours extracted from the stimuli (see Appendix A, figure 5), not all stimuli follow the dominant pitch contour pattern. This variation is to be expected since these calls are not learned signature songs. For crow *A*, the pitch contour of stimuli 2 and 3 resemble the most frequent pitch contour of crow *B* (constant low), and for crow *B*, stimulus 1 resembles the most frequent pitch contour of crow *A* (low to high). Stimuli a2, a3, and b1 are therefore expected to trigger more mistakes than the other stimuli. These “problematic” stimuli that trigger the most mistakes will be compared to the main values of b stimuli with regards to the acoustic cues mean pitch, mean F1, F1-F2 dispersion, pitch contour, intensity contour, formant contour, local shimmer, local absolute jitter, auto-correlated harmonicity, noise-harmonicity ratio, and harmonicity-noise ratio. This should provide additional insights into the relevance of the diverse acoustic cues. Jitter describes the pitch variation across cycles, shimmer describes the variation of peak-to-peak intensity (Farrús, Hernando, & Ejarque, 2007), while harmonicity and the subsequent ratios to noise describe the acoustic periodicity (Popham, Boebinger, Ellis, Kawahara, & McDermott, 2018) and how much of the sound is periodic as opposed to noise (Qi & Hillman, 1997). Jitter, shimmer, and the harmonicity values provide additional information in voice discrimination and recognition tasks and were obtained for crow *B* with the voice report function in Praat (Boersma & Weenink, 2019, see Appendix A, table 8) to explore all measurable acoustic features as possible discrimination cues.

The analysis of a potential learning trend was identical to experiment 2. One participant had no success streak and was counted as needing 40 trials, two were excluded due to a d’ score of 0.

The confidence interval was calculated in R based on mean (a), standard deviation (s) and sample size (n) using the formula qnorm(0.975)*s/sqrt(n) for the error and a+-error for the interval margins. Cohen’s d for the one-sample Wilcoxon signed-rank test was calculated with formula 1 (see experiment 1). Cohen’s for the Mann-Whitney U test was calculated with formulae 2 and 3 (see experiment 2). Power was calculated with GPower (Erdfelder et al., 1996) version 3.1.

4.2. Results

The average d’ score was 2.48 (SD=1.1, 95% CI [2.17, 2.78]) and individual scores ranged from 0 to 5.13. D’ scores were significantly above chance level (one-sample t-test with mu of 0, t=15.86, df=49, d=2.25, power=1, p<0.01). A Mann-Whitney U test revealed that d’ scores differed significantly both between experiment 1 and 3 (W=669, d=0.9, power=0.99, p<0.01), as well as experiment 2 and 3 (W=1024.5, p<0.01, d=1.22, power=0.99; see figure 18).

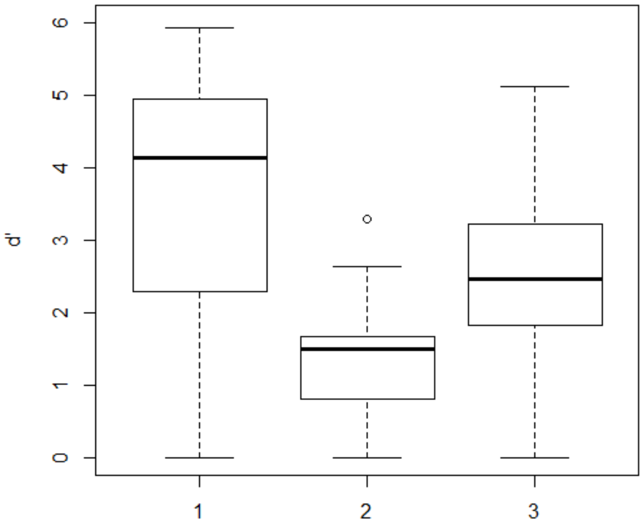


Figure 18: d’ scores compared between experiment 1 (zebra finch song with pitch contour), experiment 2 (zebra finch song without pitch contour) and experiment 3 (crow calls), created in R version 3.6.1 (R Core Team, 2019). Mean d’ scores are 3.68, 1.3, and 2.48 respectively.

A Welch t-test showed no significant difference in d’ scores between female and male participants (t=0.57, df=22.1; see figure 19).

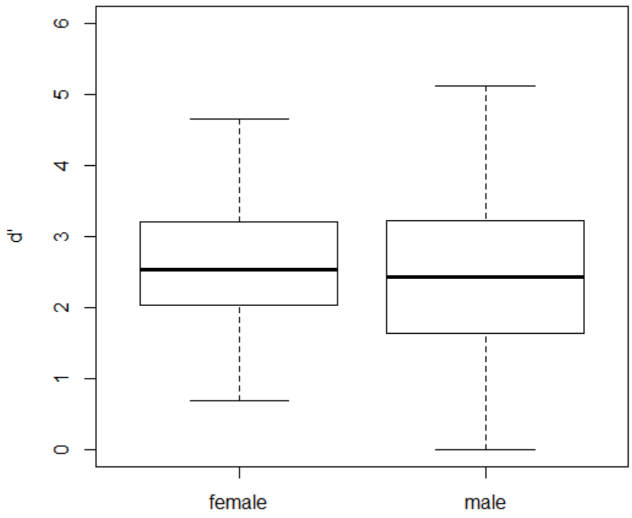


Figure 19: d’ scores obtained by 32 female and 17 male participants, created in R version 3.6.1 (R Core Team, 2019). Mean d’ scores are 2.59 and 2.37 respectively.

A Welch t-test also found no significant difference in d’ scores between participants with and without a musical background (t=1.17, df=21.1; see figure 20).

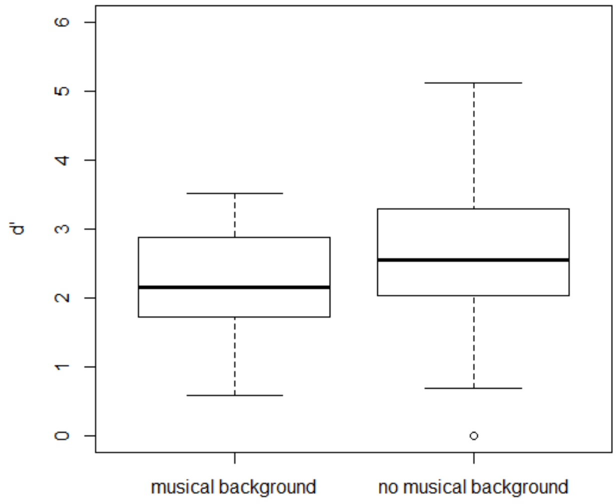
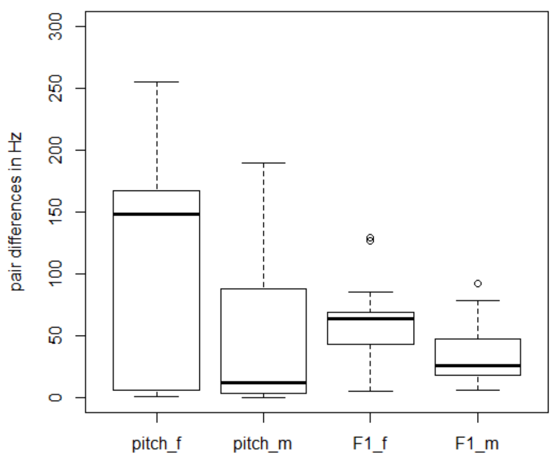


Figure 20: d’ scores obtained by 11 participants with a music background and 39 participants without a musical background, created in R version 3.6.1 (R Core Team, 2019). Mean d’ scores are 2.19 and 2.56 respectively.

Nine “miss” responses out of 265 were triggered by pairs with identical stimuli (e.g. a6a6). A Mann-Whitney-U test did not show a significant difference between pitch differences that triggered “false alarm” responses and those that triggered “miss” responses (W=245.5 p=0.0506), but did show a significant difference for F1 differences (W=262, p<0.05) and formant dispersion differences (262, p<0.01, see figures 21 and 22).



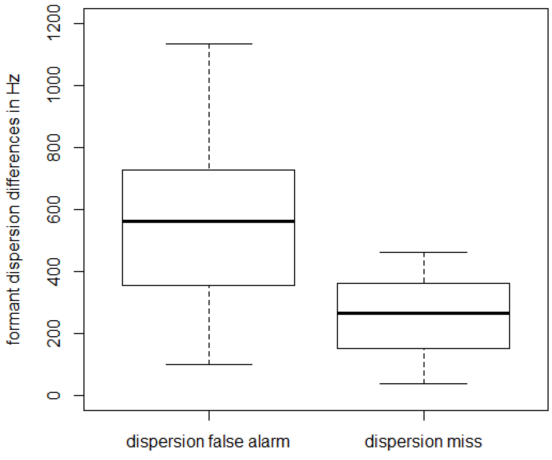
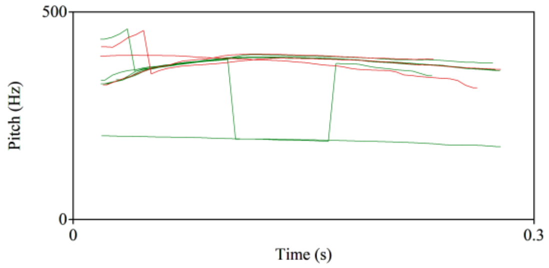


Figure 21: difference in pitch and F1 frequency of each stimulus pair in Hz that triggered the same response (false alarm, miss) at least five times, created in R version 3.6.1 (R Core Team, 2019). Mean pitch differences were 111Hz, 48Hz, 60Hz, and 34Hz respectively.

Figure 22: difference in F1-F2 dispersion difference of each stimulus pair in Hz that triggered the same response (false alarm, miss) most often, created in R version 3.6.1 (R Core Team, 2019). Mean dispersion differences were 574Hz and 245Hz respectively.

The stimuli that appeared most often in incorrectly categorized pairs were a3 (66 pairs), a4 (64 pairs), and a5 (68 pairs), as well as b4 (87 pairs), b2 (85 pairs), and b8 (74 pairs; see Appendix A, table 7). The stimuli pairs b2b4, b8b4, and b4b6 also triggered the response "miss" unusually often (17, 11, and 10 times respectively). A comparison of pitch, F1, F1-F2 dispersion, local shimmer, local absolute jitter, harmonicity, noise-harmonicity ratio, and harmonicity-noise ratio between the problematic stimuli and the mean values for the b stimuli provided no explanation for this pattern (see Appendix A, table 8 for full report). A comparison of pitch contour, intensity contour, and formant contour between the problematic and the unproblematic b stimuli also showed no visible difference between the two (see figures 23-25).



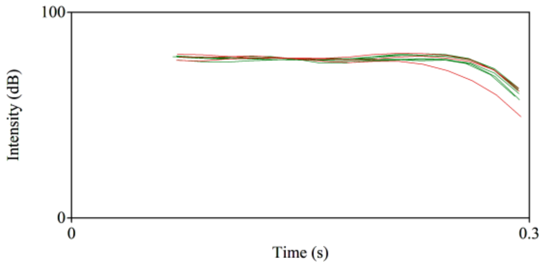


Figure 23 (left): Pitch contour of the problematic stimuli in red (b2, b4, and b8) compared to the unproblematic stimuli in green (b1, b3, b5, b6, b7, b9), created in Praat version 6.0.49 (Boersma & Weenink, 2019).

Figure 24 (right): Intensity contour of the problematic stimuli in red (b2, b4, and b8) compared to the unproblematic stimuli in green (b1, b3, b5, b6, b7, b9), created in Praat version 6.0.49 (Boersma & Weenink, 2019).

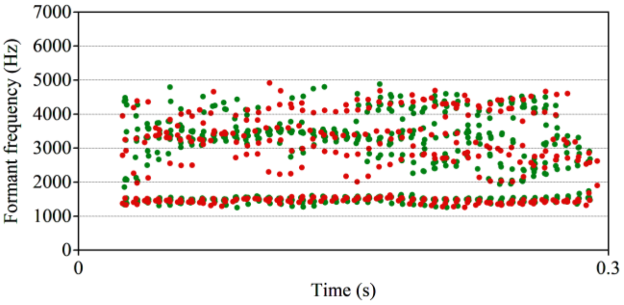
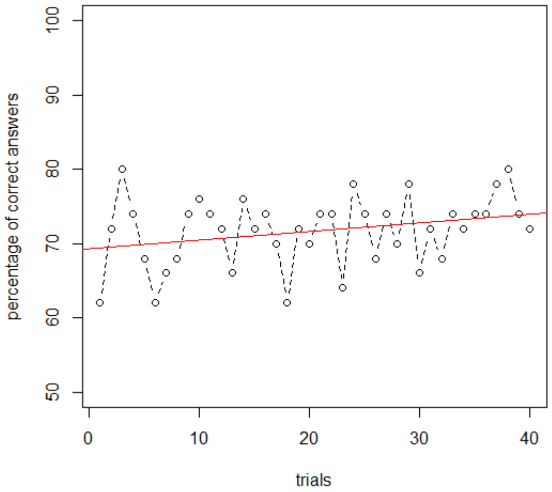


Figure 25: Formant contour consisting of the formants F1, F2, and F3 of the problematic stimuli in red (b2, b4, and b8) compared to the unproblematic stimuli in green (b1, b3, b5, b6, b7, b9), created in Praat version 6.0.49 (Boersma & Weenink, 2019).

The average participant needed 11 trials (SD=10.56) before reaching the first success streak (threshold of 5 correct answers in a row) and made 11 mistakes (SD=4.59) in total. A Mann Kendall Trend test showed no significant, although positive trend across all trials (τ =0.19; see figure 26), as well as no significant trend across the first 20 trials (τ =0; see figure 27).



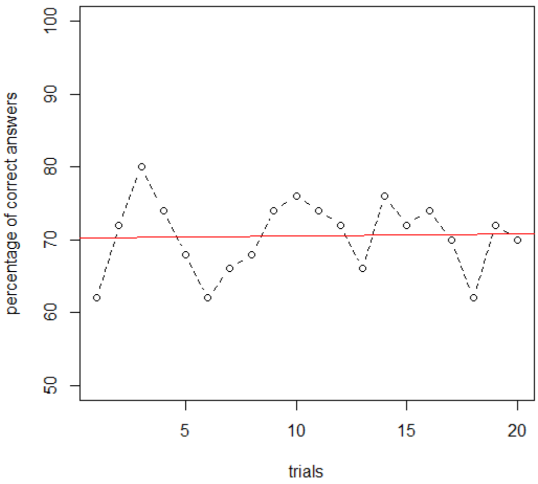


Figure 26 (left): Percentage of correct answers (either “hit” or “correct reject”) for each trial with one answer per participant and trial and a linear regression line indicating overall trend, created in R version 3.6.1 (R Core Team, 2019).

Figure 27 (right): Percentage of correct answers (either “hit” or “correct reject”) for each of the first 20 trial with one answer per participant and trial and a linear regression line indicating overall trend, created in R version 3.6.1 (R Core Team, 2019).

5. Discussion

These results show that human adults can discriminate two individual zebra finches, as well as two individual large-billed crows based on a very short section of their respective vocalizations. This can be accomplished with and without a signature vocalization as additional identifier, although performance was highest for the zebra finch signature songs.

Neither participants' sex nor their musical background had an effect on the discrimination success. This does not match previous findings that people with musical background are more sensitive to pitch differences and global voice timbre than non-musicians (Chartrand & Belin, 2006; Pitt, 1994), or that men perform better in acoustic size judgement tasks due to a higher sensitivity to formant dispersion (Charlton et al., 2013), suggesting that participants relied on different acoustic cues in this discrimination task. It should however be noted that the extent of individual musical backgrounds is not known and might not have crossed a necessary threshold in some or all participants. Additionally, musical background was self-reported by participants and the criteria necessary for this classification may also vary between individuals. As the number of participants with musical background was also very low compared to those without it, the analysis of this characteristic should only be taken as cautious suggestion and not as an absolute rejection of musical experience as a contributor to the discrimination accuracy.

While the analysis of the correlation between frequency differences of multiple acoustic cues within pairs and the response type they triggered was partially significant, none of these results support that these cues were used in the discrimination, as the correlation between response type and frequency difference is the opposite of what was expected. "False alarm" responses showed larger differences in F1 and formant dispersion frequencies when small differences were expected (since small differences would be more likely mistaken as "same individual"), and "miss" responses showed small differences when large once were expected (since large differences would be more likely mistaken as "different individual"). In other words, I would have expected the opposite of the relative proportions in "false alarm" and "miss" responses shown in figures 21 and 22. Mean pitch differences, for example, would have been expected to be significantly lower in "false alarm" responses than in "miss" responses, but the mean differences observed were 382Hz to 309Hz for zebra finch songs, and 111Hz to 48Hz for crow calls. These mean pitch differences for "false alarm" responses are also far above the minimum perceivable frequency difference of 0.2–0.3% for frequencies between 250–4000 Hz (Moore 1973, as cited in Lopez-Poveda, 2014) which would translate to a perceivable pitch difference of 5-10Hz with these stimuli. As such, the analysis only showed that mean F1 and formant dispersion varied little intra-individually while differing more inter-individually for both natural zebra finch songs and crow calls, while pitch varied considerably intra-individually in the zebra finch songs. In addition to that, multiple pairs containing identical stimuli (e.g. a2a2) triggered "miss" responses, meaning that participants did not judge them to be made by the same individual despite being absolutely identical in every acoustic feature. Neither mean pitch, nor mean F1 or the dispersion between F1 and F2 appear to have played a role in any of the three discrimination tasks. Since these cues were not used for the discrimination, it is not surprising that neither male participants nor participants with musical backgrounds obtained significantly higher d' scores.

The decline in discrimination accuracy between experiments 1 and 2 and the highly stereotyped pitch contour patterns in both individuals suggest that pitch contour was an important acoustic cue in the discrimination of individual zebra finches. It is however not the only relevant feature, as participants were still able to discriminate the stimuli without pitch contour. A previous study found no significant effect of musical experience on the discrimination of pitch contour prior to training (Wayland, Herrera, & Kaan, 2010), and so participants with musical background would not be expected to perform significantly better than other participants. This does however not seem to be the case for the discrimination of individual crows. The analysis of stimuli occurrence in incorrectly categorized pairs in experiment 3 did not show a correlation between atypical pitch contour (pitch contour that resembles the typical pitch contour of the other crow) and a higher occurrence in incorrectly categorized pairs. Out of the six stimuli with the highest occurrence, only a3 shows an atypical pitch contour. The direct comparison of the highly problematic stimuli b2, b4, and b8 also showed no difference in pitch contour, intensity contour, formant contour or any of the measured acoustic values compared to the mean values from crow *B*. There is currently no explanation as to why these three stimuli over-proportionally triggered incorrect responses. Pitch contour does therefore not appear to have been a relevant cue in the discrimination of individual crows, as opposed to the discrimination of individual zebra finches. Whether this might be due to species bias (e.g. the crow calls being perceived as less melodic or less pleasant), the pitch contour in crow calls being less stereotyped or something else entirely cannot be said at this point.

It should however be noted that these cues were analysed at a group level. It may be the case that cue relevance and perceptual strategies vary between individuals due to selectively paid attention to different acoustic dimensions (Holt et al., 2018). Different strategies may lead to contradictory patterns between participants, such as the stimuli pairs that triggered opposing responses (e.g. both "false alarm" and "miss"). Additionally, the analyses undertaken here only take suprasegmental cues into consideration, e.g. mean pitch and formant frequencies for the entire song segment. It is possible that participants' discrimination varies due to individual differences in perception, as well as attentional biases based on previous linguistic (such as acquisition of a tonal language that encodes meaning through pitch, see Chang, Yao, & Huang, 2017) and non-linguistic (such as musical background) experience.

No significant continuous learning effect was found, although a rapid increase in correct answers can be seen between the first and second trial in experiment 1, suggesting that participants need at least one example of stimuli from both individuals before they are able to reliably discriminate them (assuming that most participants have heard at least one sound segment of both birds by the end of the second trial). This is not a sign of learning, but rather the consequence of not having any information about the sounds to be compared other than them belonging to individuals of the same (unnamed) species. If I had asked them whether an orange is like an apple the right answer would have depended entirely on whether the categories I had in mind were "red" and "orange", or "fruit" and "furniture". There is a small upwards trend in the first 20 trials in experiment 1, although not in experiments 2 (decline) and 3 (level). There is also slight, albeit non-significant upwards trend across all trials in experiment 3 suggesting a learning effect in the second half of the experiment. If this task was purely based on learned discrimination, I would have expected to see a success percentage of roughly 50% (chance level) on the first trial, followed by an increase and a plateau at a percentage above the initial 50%. For the three experiments, the percentages of correct answers on the first trial were however 70%, 68%, and 62% respectively. While there may have been individual improvement of varying degrees, the initial success rates above chance level suggest that learning during the task is not a necessity for discrimination. It is however unclear to what extent previously learned skills can support the discrimination task. The slight decline across all trials in experiments 1 and 2 points to a drop in performance due to fatigue (judging by some participants asking how much longer it will take), lack of reinforcement, or increasingly overlapping memories of stimuli introduced by the high number of trial repetitions. It may be advisable to reduce the number of trials in future experiments to avoid this effect. The steeper decline observed in the group of participants who scored below average compared to the entire sample in experiment 1 suggests that lack of motivation played a role in their below-average performance. It should however be noted that all trends were analysed on the group level, not the individual level. Based on the great variance of number of trials needed before participants reached the five-in-a-row success streak, it is possible that some participants do improve to some degree. However, quantifying individual improvement with the binary data points (correct vs incorrect) for each person would be statistically unreliable. A future study might explore potential long-term learning across multiple data collection sessions to examine consistent improvement and individual differences. Together with the genetic underpinnings of acoustic perception presented in the introduction, the above-chance level first trial success percentages, the absence of an overall significant learning trend, the high inter-individual variation discussed in the introduction, and the non-effect of a musical background (which is considered a prior experience) suggest that the discrimination ability is at least partially based on a genetically determined ability, although experience other than musical background may contribute to some (currently unknown) degree as well. Again, I would like to stress that this ability is not exclusively used for individual heterospecific discrimination. There is no reason to believe that a human sensitivity to individuality in zebra finches or crows would be selected for (see Chapter 2 and 3). It is more likely based on one or multiple perceptual abilities that are sensitive to some acoustic cues that are also present in these vocalizations, such as pitch contour, and can therefore be used for the discrimination of pitch contour in any acoustic format, heterospecific amniote or other.

The d’ scores reported here for both zebra finches and crows are considerably higher than those observed for rhesus monkeys and equal to those observed for human voices and crow calls by (Friendly et al., 2014). This may be due to methodological differences as the previous study used sound pairs matched for mean, minimum, and maximum pitch instead of randomizing stimuli pairings. It is also possible that other factors that were not considered influenced individual discrimination performances, such as above average exposure to birdsong (e.g. growing up in a rural environment, personal interest in bird song etc.), familiarity with a tonal language (such as tonal languages), or difficulty understanding the task. While all participants verbally confirmed having understood the task, some seemed to find it difficult to understand that individual birds of the same species could sound different, which is a bias that may have influenced their performance. Future experiments could explore the potential relevance of temporal cues and timbre in the discrimination task, as well as a possible correlation between response time and certain stimuli pairs (e.g. consistently shorter for pair a1b1 and consistently longer for a2b2), which may provide further insights into the relevance of the various acoustic cues.

Chapter 2: Function

1. Introduction

1.1. Inclusive Fitness

Evolution is the change of allele frequency within a population and between generations (Futuyma, 2013). Selective pressures, both biotic and abiotic, lead to nonrandom differences in survival between individuals. As such, strategies that are best adapted to the given environment are favoured through increased reproductive success, which in turn leads to the increased or decreased representation of a given genotype in gene pool of the the next generation. This is called Darwinian Fitness, which everyone in the population seeks to maximize for themselves. In other words, from an evolutionary point of view an individual's goal is to have its genotype represented as much as possible in the next generation's gene pool (Darwin 1859, as cited in Gardner, 2017). An individual's fitness does however not only depend on its own reproductive success, but also on that of close relatives. Diploid organisms receive 50% of their genetic material from each parent (i.e. one set of chromosomes from each parent), giving a relatedness of 50% between offspring and parent, or a coefficient of relationship of 0.5. This means that at any locus (location of a gene), the likelihood of parent and offspring having the same allele (variant of a gene) is 50%. Since the individual's siblings also receive 50% from each of the same two parents, full siblings have a coefficient of relationship of 0.5 as well (Wright, 1922). For the goal of increasing one's fitness (share of own genotype in the gene pool of the population), it doesn't matter whether the allele in question comes from the individual themselves or from a relative who has the same allele. Selection may therefore act on an individual through their kin with whom they share a certain proportion of their genotype. This process is termed kin selection, termed by Hamilton. Because of the difference in proportion of shared alleles depending on the coefficient of relationship, the kin selection likelihood depends on the degree of relatedness. This is measured by Hamilton's rule, rb-c>0 (relatedness at the loci of interest*benefits-costs>0; Hamilton, 1964). A genotype's contribution to the gene pool over multiple generations then determines the individual's inclusive fitness, based on the number of their own viable and in turn successfully reproducing offspring, as well as their relatives' offspring.

In the following, I will discuss the function and inclusive fitness consequences of individual discrimination and recognition for the purposes of kin selection, social groups and cooperation, pair bonds, parental care, territoriality, co-habitancy with heterospecifics, followed by potential costs incurred from a recognizable voice. While individual discrimination for purposes such as kin selection or parental care are only relevant between conspecifics but not heterospecifics, these examples are still included because heterospecific discrimination could rely on the same perceptual abilities as conspecific discrimination. Although animals may just as well vocalize without intention (spontaneous signalling), I will focus on vocalizations directed at a recipient.

1.2. Functions of Acoustics Individual Discrimination and Recognition

Since kin share considerable proportions of their genotypes, choosing a closely related mate leads to offspring with decreased genetic variation and low fitness, also referred to as inbreeding depression (Bouzat, 2010). Kin discrimination and recognition is therefore an important ability for individuals to increase their indirect fitness through cooperative behaviour directed at relatives, and their direct fitness through reproductive success, especially since related individuals tend to live in close vicinity to each other due to limited dispersal (Tibbetts & Dale, 2007). Long-tailed tits (*Aegithalos caudatus*) discriminate between kin and non-kin by their contact calls (Sharp, McGowan, Wood, & Hatchwell, 2005), and female zebra finches recognize the song of their father (Miller, 1979).

Some species may live in social groups, in which case information transfer between group members becomes relevant as well. Zebra finches live in colonies of up to 350 individuals where they participate in social activities in a “social tree” and also remain in almost constant contact with their mate (Swaddle, 2010; Zann, 1996). Large-billed crows live in fission-fusion societies (Kondo, Izawa, & Watanabe, 2010), which is a social system that flexibly increases and decreases its number of sub-groups based on available resources (Kummer 1971, as cited in Aureli et al., 2008). Unlike most songbirds, crows don’t sing to defend their territory or initiate copulation, but rather for these intergroup bonding purposes (Brown, 1985). Large-billed crows recognize individual members of their group by their contact calls (Kondo, Izawa, & Watanabe, 2012). Kondo, Izawa and Watanabe (2010) also found that the ka-calls of large-billed crows produced by two individuals in a social context follow a temporal pattern of set interval lengths between calls, as opposed to individually timed intervals in groups with multiple individuals. This temporal rule is comparable to turn-taking, making vocal communication more efficient. Izawa and Watanabe (2008) found a linear dominance hierarchy in hand-raised, captive large-billed crows in which transitive dominance relationships remain stable over time, thus requiring individual recognition. In these hierarchically organized flocks, individuals promote their social status through the number of sequential-note calls they produce, and dominance is correlated with a higher frequency of sequential-note calls but not contact calls (Kondo & Hiraiwa-Hasegawa, 2015). Group members may also cooperate with each other. Cooperation can be divided into mutual benefit (mutualism), which has a positive fitness consequence for both the actor and the recipient, and altruism, which has a positive fitness consequence for the recipient but comes at a cost for the actor (Hamilton, 1964). In unrelated individuals where kin selection does not apply, this is only expected in the form of reciprocal altruism where the sender expects the recipient to return the favour at a later time (Trivers, 1971). This will be discussed in detail in the next section. African elephants (*Loxodonta africana*) discriminate the infrasonic contact calls of female family and group members from those of non-group members (McComb, Moss, Sayialel, & Baker, 2000), domestic horses (*Equus caballus*) cross-modally recognize herd members using auditory and visual/olfactory cues (Proops, McComb, & Reby, 2009), and greater mouse-eared bats (*Myotis myotis*) use formant-related features in echolocation calls to discriminate conspecifics within their group (Yovel, Melcon, Franz, Denzinger, & Schnitzler, 2009). In humans, the accuracy of voice recognition is correlated with the perceived familiarity with the respective speaker (Wenndt, 2016).

In some species individuals form pair-bonds to breed, either for a limited period of time or for their entire lives. This is often found in biparental species where it is advantageous to retain the same breeding partner over multiple breeding seasons (Forslund & Larsson, 1991; Wittenberger & Tilson, 1980). These pair-bonds can only be maintained if an individual can accurately discriminate their mate from a stranger. Zebra finches form life-long bonds, and females can discriminate between their mate’s song and that of a neighbouring male (Miller, 1979). In addition to recognition based on signature songs, pairs also recognize each other by the structure of their contact calls (Vignal, Mathevon, & Mottin, 2004, 2008). Large-billed crows also form life-long, monogamous pair-bonds (Matsubara, 2007). Although crows mostly produce innate calls, both sexes also sing songs throughout the year. These songs are shared within their mate pair, include elements that are tied to social interactions within their group and dependent on the relationship between individual group members. (Brown, 1985; Goodwin, 1976). When in a crowd with competing speakers, humans can use the high degree of familiarity with their spouse to both selectively focus on and selectively ignore their voice, depending on the conversation requirements (Johnsrude et al., 2013).

Parental care is an important factor in the reproductive success of some species, as their offspring would not survive without the resources provided by a parent. While it is advantageous for individuals to provide parental care for their own offspring, investing resources in unrelated offspring, such as the offspring of brood parasites (Davies & Brooke, 1989) would be disadvantageous. The discrimination of offspring and non-offspring is therefore an important ability to maximize parental care efficiency and thus reproductive fitness. In turn, offspring profit from discriminating their parents from strangers who do not provide care, a skill that is relevant for chicks in colonial breeding species, such as king penguins (*Aptenodytes patagonicus*), who have to localize their parents in a crowd of conspecifics in order to be fed (Jouventin, Aubin, & Lengagne, 1999). Domestic cattle (*Bos taurus*), domestic sheep (*Ovis aries*), and domestic pigs (*sus scrofa domestica*) show mother–offspring recognition using contact calls (Illmann, Špinka, Schrader, & Šustr, 2002; La Padilla de Torre, Briefer, Ochocki, McElligott, & Reader, 2016; Searby & Jouventin, 2003), young Australian sea lions (Neophoca cinerea) still recognize their mother's voice two years after weaning (Pitcher, Harcourt, & Charrier, 2010) and young domestic chickens (*Gallus gallus domesticus*) can learn to discriminate the parental call of individual hens when paired with a visual stimulus during training (Cowan, 1974). Crocodile juveniles emit pre-hatching calls in the egg to synchronize hatching and to stimulate the mother to unearth the eggs (Vergne & Mathevon, 2008). Zebra finch pairs use structured call duets to coordinate parental care during incubation (Boucaud, Mariette, Villain, & Vignal, 2016). In crows, both sexes contribute to nest building and parental care (Goodwin, 1976) and males will provide food to their mate (Matsubara, 2007). If nesting crows think that their offspring is threatened, they will produce scolding calls in their defense (Good 1952 as cited in (Chamberlain & Cornwell, 1971). While still in utero, human foetuses show an increased heart rate when hearing their mothers voice, as well as a decreased heart rate when listening to the voice of a stranger (Kisilevsky et al., 2003). Less than 24 hours after birth, neonates recognize their parents' voices and show a decreased heart rate when listening to them, compared to an acceleration when listening to unfamiliar voices (Ockleford, Vince, Layton, & Reader, 1988).

Individuals of some species inhabit territories, and defending this territory against rivals is a vital part of gaining and maintaining access to potential mates. But fights are also costly and should be avoided when possible. The discrimination of familiar neighbours and unfamiliar rivals cuts down on unnecessary fights (Tibbetts & Dale, 2007), which is also called dear enemy effect (Fisher, 1954). Male concave-eared torrent frogs (Odorrana tormota), agile frogs (*Rana dalmatina*) and male white-throated sparrows (Zonotrichia albicollis) discriminate neighbours from strangers by their vocalizations (Brooks & Falls, 1975; Feng et al., 2009; Lesbarrères & Lodé, 2002). Both zebra finches and crows produce calls to defend their territory (Swaddle, 2010; Tarter, 2008).

In addition to conspecifics, many animals are also regularly in contact with members of other species due to shared habitats. The discrimination of heterospecifics thus further extends the above-listed benefits of discriminating between familiar and unfamiliar, potentially dangerous individuals. Heterospecifics may exchange information, behave cooperatively towards another, and even form heterospecific social groups (Sridhar & Guttal, 2018; Stahler, Heinrich, & Smith, 2002; Tibbetts & Dale, 2007). Carrion crows (*Corvus corone*) and captive cheetah (*Acinonyx jubatus*) discriminate between familiar and unfamiliar human voices (Leroux, Hetem, Hausberger, & Lemasson, 2018; Wascher, Szipl, Boeckle, & Wilkinson, 2012), domestic dogs and domestic cats of multiple breeds discriminate their owner's voice from that of an unfamiliar person (Adachi, Kuwahata, & Fujita, 2007; Saito & Shinozuka, 2013), rhesus monkeys can match a familiar human voice to the corresponding face (Sliwa, Duhamel, Pascalis, & Wirth, 2011), and Campbell’s monkeys (*Cercopithecus campbelli*), Guereza colobus monkeys (*Colobus guereza*), and redcapped mangabeys (*Cercocebus torquatus*) can discriminate between familiar and unfamiliar De Brazza monkeys (*Cercopithecus neglectus*) by their contact calls (Candiotti, Zuberbühler, & Lemasson, 2013). An additional advantage of attention to heterospecific vocalizations includes the exploitation of alarm calls. Oriental reed warblers (*Acrocephalus orientalis*) and black-browed reed warblers (*Acrocephalus bistrigiceps*) eavesdrop on each other’s alarm calls with regards to cuckoo sightings, a brood parasite that targets both species (J. Yu et al., 2019). Eavesdropping on the vocalizations of other species has also been observed in superb fairy-wrens (*Malurus cyaneus*; Magrath, Haff, McLachlan, & Igic, 2015), non-vocal Galápagos marine iguana (*Amblyrhynchus cristatus*; Vitousek, Adelman, Gregory, & Clair, 2007), non-vocal Madagascan spiny-tailed iguana (*Oplurus cuvieri cuvieri*; Ito & Mori, 2010), nonvocal, white-bellied copper-striped skinks (*Emoia cyanura*; Fuong, Keeley, Bulut, & Blumstein, 2014), yellow-bellied marmots (*Marmota flaviventris*) and golden-mantled ground squirrels (*Spermophilus lateralis*; Shriner, 1998), collared pika (*Ochotona collaris*; Trefry & Hik, 2009), non-social Gunther's dik-diks (*Madoqua guentheri*; Lea, Barrera, Tom, & Blumstein, 2008), zebra finches (Guillette, Hoeschele, Hahn, & Sturdy, 2013), black capuchin monkeys (*Sapajus nigritus*; Wheeler, Fahy, & Tiddi, 2019), ashy-headed laughingthrushes (*Garrulax cinereifrons*), orange-billed babbler (*Turdoides rufescens*; Goodale & Kotagama, 2008) and bonnet macaques (*Macaca radiata*; Ramakrishnan & Coss, 2000), to name just a few. Wild-caught urban large-billed crows are less attentive to playback of the familiar language Japanese than the unfamiliar language Dutch, possibly due to eavesdropping on human speech (Schalz & Izawa, 2020).

In addition to the above-listed payoffs, being recognizable also includes costs for the signal sender. In species that do not have completely monogamous pair bonds, both sexes of a given species are in a constant arms-race against each other (sexual conflict) to increase their own reproductive success, which in turn diminishes that of the other due to sexually antagonistic phenotypes (Perry & Rowe, 2015). Briefly, both sexes receive fitness benefits from mating with multiple mates, but it is costly for them if their partner does the same. For this reason, both sexes need to rely on deception to maximise their own reproductive success at the cost of the other (Petrie & Kempenaers, 1998). This deception becomes increasingly difficult through individualized vocalizations that increase the conspicuousness of extra-pair copulation, especially in species that rely on their vocalizations to attract a mate (Tibbetts & Dale, 2007). Recognizability also decreases the chances of success for a cheater invading a social group to profit from their cooperative behaviour without returning the favour, forcing this individual to spend resources on reciprocal behaviour instead of only receiving them (Soberon Mainero & Martinez del Rio, 1985). Sensitivity to acoustic cues can also make an individual an easy target if their vocalization conveys small body size (Reby, McComb, Darwin, & Fitch, 2005). A cost for the signal receiver would be the risk of deception by a signal sender. Fork-tailed drongos (*Dicrurus adsimilis*) produce false alarm calls mimicking species-typical warning calls of other birds or mammals and then steal their food while its victims hurry away to safety (Flower, Gribble, & Ridley, 2014). Bottlenose dolphins (*Tursiops truncatus*) eavesdrop on the advertisement calls of their prey, gulf toadfish (*Opsanus beta*) to track them, while the toadfish in turn eavesdrop on low-frequency sounds produced by the dolphins but ignore low-frequency sounds produced by shrimp to reduce advertisement calls in the presence of a predator (Remage-Healey, Nowacek, & Bass, 2006). Recognizability is therefore a disadvantage for both species that each has to counter-act with an improvement of their own recognition abilities.

The discrimination and recognition of individuals, both conspecifics and heterospecifics, may thus provide a number of benefits, including cooperation. As cooperating with the wrong individuals would lead to unnecessary fitness costs, natural selection should favour reliable recognition of individuals, including recognition by voice. The following section thus discusses reciprocal altruism, one form of cooperation, as well as the necessary recognition abilities to link an individual’s cooperation partners’ reputation and past experiences with their voice as a means of identification.

1.2.1. Reciprocal Altruism

Reciprocal altruism was first described by Trivers (1971). In his seminal paper, he establishes how altruism between unrelated organisms is favoured under natural selection if the favour provided to B by A will later be reciprocated by B, and assuming that the benefit to B (measured as fitness increase) is greater than the cost to A (measured as fitness decrease). Cheating (defined as “failure to reciprocate”), on the other hand, would not be favoured by natural selection if the costs are greater than the benefits, for example id A refuses to provide any future favours. This theory has been summarized by (Stephens, 1996) into the following five (theoretical) conditions:

1. To the donor, cooperation must be more costly than selfish behaviour in order to be altruistic.
2. To the receiver, the donor’s cooperation must be more beneficial than the donor’s selfish behaviour in order to be altruistic.
    a. In the strong sense, this benefit is independent of the receiver’s own behaviour (cooperation or defecting).
    b. In the weak sense, this benefit is only present if the receiver defects.
3. The above conditions must apply to both parties in order to be reciprocal.

4. The cooperation cannot be contingent on the expectation of an immediate benefit as that would be mutualism.
5. In order for the cooperation to evolve, there must either be cheating detection or infinite iterations (i.e. the two parties do not know which cooperation event will be the last one).

Some might argue that the term reciprocal altruism is self-contradictory because altruistic behaviour has to be costly to the donor, but eventual reciprocation offsets the initial cost. This leaves both parties with received benefits and would therefore qualify as mutualism, or reciprocal cooperation (Axelrod & Hamilton, 1981). This has been the subject of an ongoing debate, with critics arguing that reciprocal altruism is not truly altruistic since "it provides a direct fitness advantage to cooperat[e]. If an individual does not pay the cost of cooperation in the short term then it will not gain the benefit of cooperation in the long term" (West, Griffin, & Gardner, 2006). But the fitness advantage is not actually direct because it is contingent on reciprocity at a later time, which is not guaranteed, a detail that has been established in condition 4 (see above). At the time of cooperation, neither the donor nor the receiver can know whether the receiver will ever be willing or able to reciprocate. For instance, vampire bats (*Desmodus rotundus*) have been found to donate blood to starving conspecifics on a reciprocal basis (Wilkinson, 1984). However, each donation event consists of the donor paying a cost while only the receiver of the blood receives a fitness advantage. It is in no way guaranteed that the receiver will ever have blood to share, or whether the donor will ever need a donation. This is in contrast to examples of mutualism where both parties receive a benefit from a single cooperation event. In a mutualistic foraging partnership, honeyguides (*Indicator indicator*) lead humans to bee colonies where they will then collect the honey and leave the empty hives from which the honeyguides can eat the larvae and wax (Spottiswoode, Begg, & Begg, 2017). Furthermore, the argument proposed by West et al. (2006) does not take into account that cheating detection is imperfect and cheaters may exploit cooperation without being detected (see below), in which case they may well continue to receive the benefit of cooperation without paying the costs. Considering this uncertainty surrounding delayed reciprocation, such a cooperation is at least temporarily altruistic until it is actually reciprocated when the donor becomes the receiver, a detail that is best captured by Triver's term "reciprocal altruism".

An individual may base its cooperation decisions on certain strategies dictating when to cooperate and when to defect. An Evolutionary Stable Strategy (ESS) is a strategy (or phenotype) that, if assumed by the entire population, cannot be invaded by a mutant strategy. A mutant strategy is an initially rare alternative to the strategy being played by the population. Whether or not a strategy is an ESS depends on the circumstances around the cooperation situation. For instance, a population consisting entirely of co-operators can be invaded by a cheater who receives the benefits of cooperation but does not reciprocate. A population entirely free of cooperators cannot be invaded by a cooperator, as there would be no one to cooperate with (Maynard Smith & Price 1973, as cited in Maynard Smith, 1982). This can be modelled in a variety of games following different sets of rules and scenarios. One often referenced game is the Prisoner's Dilemma, which was conceived by Flood & Dresher (1950, as cited in Mérő, 1998) and further developed by Tucker (1951, as cited in Mérő, 1998). In this game, two prisoners housed in separate cells are asked to testify against each other, in which case they themselves will be set free and the other prisoner will serve ten years. If both confess, both will be incarcerated for five years and if neither confesses, both only serve one year (see table 4 for payoff matrix; Mérő, 1998). In this setup, mutual silence corresponds to cooperation, while confessing is considered defecting (one-sided confessions being comparable to cheating). In the iterated Prisoner's Dilemma where multiple rounds are being played, prisoners can take each other's previous decisions into account as well. According to Axelrod and Hamilton (1981), the best solution to both the one-shot Prisoner's Dilemma and the iterated Prisoner's Dilemma with a known number of rounds is always to defect, making this strategy an ESS. This is because defecting is a strategy that cannot be invaded by a mutant. If A confesses, he will either get 5 years or 0 years in prison. Cooperation however depends on mutual silence and if that strategy is invaded by a mutant (not cooperating), B defects and A gets 10 years in prison, which is worse than the maximum of 5 years had he defected as well. If the number of rounds is unknown and the next round is dependent on a probability w, Axelrod and Hamilton consider both defecting and "tit for tat" (initial cooperation, followed by whichever strategy the opponent uses) to be an ESS, although the latter depends on a sufficiently high w. This is because when a next round can be expected, retaliation by the other party following defecting can also be expected and should be avoided by both parties. Tit for tat is an easier alternative to the long-term tracking of an individual's social behaviour and is preferentially played by rats (Schweinfurth & Taborsky, 2020). A higher payoff or a lower cost constitutes a fitness benefit. Natural selection should therefore favour those who play an ESS, i.e. those genotypes which code for an ESS phenotype.

Table 4: Payoff matrix for the Prisoner's Dilemma game as shown Mérő (1998). Costs (in this case years of freedom lost) to A are indicated on the left side of the slash, costs to B on the right.

	***B* confesses**	***B* does not confess**
***A* confesses**	-5 / -5	0 / -10
***A* does not confess**	-10 / 0	-1 / -1

Unlike in the one-shot Prisoner's Dilemma, repeated interactions between conspecifics are very probable within a social group. Although interactions between heterospecifics in itself would be less likely, limited dispersal, which is thought to promote altruism between kin due to physical proximity and consequential higher relatedness (Hamilton 1964 as cited in Kümmerli, Gardner, West, & Griffin, 2009), could indirectly lead to some repeated interactions between heterospecifics as well (Hamilton, 1972). As such, Trivers (1971) argues that favours provided to a local heterospecific are likely (but not guaranteed) to be returned to the donors, increasing their respective inclusive fitness and thus favouring heterospecific reciprocal altruism. One previously mentioned example is the heterospecific foraging flock of the red-browed firetails and the superb blue wrens (Forshaw & Shephard, 2012).

In the examples above, cooperation always takes place between a set number of participants A and B, which is considered direct reciprocity. Altruistic behaviour may also be reciprocated indirectly through an exchange of objects or favours between more than two parties. In this case, cooperation decisions rely on the receiver's reputation within the social group. Individual A helps individual B and in turn receives help from individual C because C knows that A has behaved cooperatively in the past. Alternatively, indirect reciprocity may also rely on a recent positive experience in that B helps C after having recently received help from A (Nowak & Sigmund, 2005). Due to the delay introduced by the proxy B, this concept is easier for cheaters to invade and thus requires individuals to keep a thorough track of others' past behaviours. This may be accomplished through policing, which reduces the fitness gain of cheating by punishing cheaters in the group (El Mouden, West, & Gardner, 2010). In human social groups, policing may range from abstract punishment in the form of virtue and societal values (Nowak & Sigmund, 2005) to concrete punishment through law enforcement. Vampire bats are able to recognize cheaters who have failed to reciprocate blood donation in the past and may refuse future donations to these individuals (Wilkinson, 1984).

In theory, the benefits of strategic cooperation may appear self-evident, but this does not mean that they are ecologically realistic and so the above-described assumptions and theories have to be empirically tested as well. To examine the cooperation abilities of individuals, several cooperation games and exchange paradigms have been developed. The loose string paradigm in which two individuals have to simultaneously pull the ends of a string to receive a reward has provided evidence for cooperation in several species, including heterospecific cooperation between dogs (*Canis familiaris*) and humans (Ostojić & Clayton, 2014). If necessary, dogs and wolves (*canis lupus*) will recruit humans as cooperation partners, and if their partner arrives late (delayed loose string task) they will wait for them before initiating the task (Range, Kassis, Taborsky, Boada, & Marshall-Pescini, 2019). Dogs also show third-party evaluation of heterospecifics' behaviour towards their human affiliates and refuse to accept food from a person who has previously defected in a cooperation task with the dog’s owner (Chijiiwa, Kuroshima, Hori, Anderson, & Fujita, 2015). Third-party evaluation has also been observed in tufted capuchin monkeys (*Cebus apella*) who are less likely to accept food from humans they have previously seen defecting in a cooperation situation (Anderson, Kuroshima, Takimoto, & Fujita, 2013). As mentioned above, humans may cooperate with honeyguides. To signal their willingness to cooperate with the honeyguides, the Mozambican honey-hunters use a special sound to which the honeyguides respond with increased probability of cooperation (Spottiswoode et al., 2017). This highlights the relevance of vocal communication in the maintenance of cooperative partnerships between heterospecifics, In a Prisoner's Dilemma game, zebra finches only consistently cooperate with their social partner, possibly because the long-term benefits of maintaining a reciprocal altruistic relationship with their mate outweigh the benefits of cheating (St-Pierre, Larose, & Dubois, 2009). Corvids cooperate spontaneously to solve various problems, including jointly pulling up the bin liner in a rubbish bin to gain access to its contents (Clayton & Emery, 2005), hunting with conspecifics (Tanalgo, Waldien, Monfort, & Hughes, 2019; Yosef & Yosef, 2010), hunting with wolves (Stahler et al., 2002), defending their territory with their mate (Bossema & Benus, 1985), and teaching others about dangerous people through vertical social learning (Cornell, Marzluff, & Pecoraro, 2012). Carrion crows discriminate between reliable and unreliable conspecifics based on their call (Wascher, Hillemann, Canestrari, & Baglione, 2015), use recruitment “kakaka” calls to attract conspecifics to a food source (Soma & Hasegawa, 2003), and accept food from familiar human feeders (Obozova, 2011 as cited in Obozova, Smirnova, & Zorina, 2018). Ravens naïve to the lose string paradigm cooperate with conspecifics to access a feeding platform by simultaneously pulling the ends of a string (Massen, Ritter, & Bugnyar, 2015). They remember the valence of their relationships with former group members for up to three years and discriminate them based on their calls (Boeckle & Bugnyar, 2012). If they are alone and struggle to access the food source on their own, they produce food calls to recruit helpers (Sierro, Loretto, Szipl, Massen, & Bugnyar, 2019). They provide long-term agnostic support for conspecifics who preen them, are related to them, or are ranked high in the dominance hierarchy. Their support of certain individuals in a conflict is likely motivated by either an already existing positive relationship or the long-term benefits of reciprocity from that individual (Fraser & Bugnyar, 2012). Ravens trained to exchange low-quality food for high-quality food with human experimenters remember whether a previously unfamiliar experimenter had treated them fairly. Both two days after the first exchange and one month after the second exchange, the ravens preferred to cooperate with the fair experimenter rather than the unfair one, showing direct reciprocal altruism between heterospecifics. Observations of the fairness of the experimenter towards a conspecific did not influence the observer’s own cooperation decisions (Müller, Massen, Bugnyar, & Osvath, 2017). Rewards also do not have to be handed out straight away:

In a delayed-exchange task with human experimenters as exchange partners, carrion crows and common ravens wait up to 320 seconds for the opportunity to exchange a previously received piece of food for a more valuable food reward instead of eating the initial one (Dufour, Wascher, Braun, Miller, & Bugnyar, 2012).

Given that these examples of cooperation between corvids and humans took place in an experimental setting with controlled pay-off, the relevance of individual recognition outside the laboratory may not be self-evident. Do humans and wild crows even cooperate autonomously, are these benefits applicable to real-world conditions? Anecdotes of people who feed crows and in turn receive various gifts suggest that cooperation between these two species does in fact occur on an individual basis (Marzluff & Angell, 2012). But even if cooperation between humans and wild crows exists, would voice recognition provide an actual benefit, or would it be redundant information? Consider the following anecdotal example: Every now and then, I go to feed a specific individual crow at the same location. The crow has habituated to this feeding ritual and approaches the feeding spot if it's nearby and sees me, as the food constitutes a benefit to the crow. There are however instances when I am at the feeding spot and the crow does not see me, in which case it misses out on this benefit. Recognition of my voice (if I were to call the crow) would reduce the number of missed feeding opportunities and this ability would thus provide a fitness benefit. Approaching any human voice would be a costly waste of energy for the crow, since not everyone will be willing to feed it. In turn, I profit from feeding the crow because interacting with wildlife and feeding wild birds has been found to provide health benefits (Cox & Gaston, 2016; Curtin, 2009) and could therefore be considered a fitness benefit to me. I cannot recognize this individual based on visual cues, and so recognizing it based on its calls would enable me to receive that benefit at locations other than the feeding spot. Approaching any crow would be a costly waste of energy for me since I am unfamiliar to these crows and they would not necessarily accept food from me. This particular example would be considered mutualism instead of reciprocal altruism, but it illustrates how recognition based on vocalizations can be beneficial for cooperation between crows and humans outside an experiment.

Considering the practicality of individual recognition for the purpose of cooperation it should now be examined whether humans are capable of recognizing an individual based on its call and link its identity to previous experiences. As already discussed in Chapter 1, voice recognition is an extension of voice discrimination and results from the discrimination experiment do not guarantee a recognition ability. It should also be noted that the main focus of this experiment is not the cooperative behaviour in itself. Any sound could be linked to an anticipated outcome and the source of said sound (in this case the crows) would not be expected to influence the participants' ability to associate a stimulus with an outcome (e.g. expectancy learning where a stimulus is “a valid predictor for the occurrence of the […] event”; Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002). The purpose is rather to examine the perceptual ability necessary to perform this task, namely to extract the relevant identifying acoustic cues from the heterospecific vocalizations, create a mental template or memory of these identifiers, and maintain them long enough to be able to link them to the predicted outcome and the anticipated payoff. Only if this is possible would it be worthwhile to consider cooperation a relevant function for the discrimination and recognition of individual vocalizations of a given species. The cooperation task here only serves the purpose of creating a context for the principal recognition – and – association task, imitating the constraints of real-world cooperation. To examine the cooperative behaviour of humans towards a heterospecific, the use of live cooperation partners (who would have to be trained to behave a certain way) would be challenging and so a virtual cooperation game would be a more feasible alternative. For this purpose, the game may be based on an exchange paradigm (object 1 of value 1 in exchange for object 2 of value 2) and exchange partners may be linked to the heterospecifics’ vocalizations for identification. In such a scenario, participants would listen to the call of one out of two (or more) crows and then be asked whether they would like to cooperate with this individual in the exchange paradigm or pass, knowing that the crow may either reciprocate or defect (see table 6 for payoff matrix).

Table 6: Payoff matrix for participants depending on their own behaviour and the crow they choose, for which the behaviour is set. Costs and benefits to the crows are not included as they are not relevant here.

	Crow *A* reciprocates	**Crow *B* defects**
Player cooperates	+1	-1
Player defects	0	0

Ideally, participants would employ a trial-and-error approach to learn which voices belong to the reciprocating individuals and which belong to the cheater, to then selectively cooperate with the former and avoid the latter. This approach would be fundamentally flawed: As mentioned above, defecting is always an ESS. By choosing to defect, participants would miss out on the benefit of object 2, but since they would also keep all of their objects 1 there would be no cost associated, as opposed to the potential costs of cheating exchange partners with the resulting loss of objects 1. In addition to the advantage of defecting to avoid costs in the current exchange, defecting would also be more sensible in the long run. To accurately predict the crows behaviour requires several rounds of learning of the crows voices and their corresponding behaviours through trial-and-error during which participants would have to gamble with their objects 1 for the possible benefit of eventually making a profit by receiving enough objects 2 to make up for the value of the lost objects 1. Looking back on the great variance in individual performance discriminating two crows in experiment 3, some participants may achieve this rather quickly while others may never reach this goal and going into the experiment, they cannot know which group they will belong to. The average participant in experiment 3 needed 11 rounds (SD=10.6) to consistently discriminate the two individuals accurately (reach a threshold of 5 correct decisions in a row, even if followed by later mistakes), and although recognition of an individual is more difficult to achieve than the discrimination of two individuals (as for the former, a mental representation has to be build, memorized and recalled) I will use this optimistic value for my payoff-prediction. If there are two different crows in this game, assuming that participants start with 20 objects 1 to trade and need to encounter both crows 11 rounds to develop a sufficiently robust mental representation of their voices, and assuming that they would have a 50% chance of being wrong in the first 22 trials (until they have heard both crows 11 times), as well as a 27% risk of being wrong in the subsequent trials (based on the data collected in experiment 3 where perfect discrimination was never achieved and the average participant made 8 mistakes in the 29 remaining rounds after round 11), they would not start to make a profit until round 23 and then still not reach a 100% profit rate (due to the 27% risk). This prediction constitutes the best-case scenario and due to the disparity of complexity between the discrimination task and the recognition task, the payoff may be delayed even further than 23 rounds. Consequently, learning to recognize the cooperative crow with the risks involved instead of defecting is not worth the small profit. This prediction also presupposes that participants employ the correct strategy throughout the entire game and look for a correlation between the crow call and the subsequent cooperation outcomes, which cannot be guaranteed or expected for participants naïve to game theory. As the likelihood of reciprocated cooperation would be 50% in this scenario, a pure cooperation strategy would on average result in a net profit of 0, with equal chances of deviations above or below this line. These arguments taken together clearly show that both in the short-run and long-run, the only reasonable strategy in this scenario would be to defect, preventing a reliable examination of the role of heterospecific voice discrimination in cooperation across species.

In the real world, the establishment of cooperation is a lot messier and is not restricted or guided by game scenario rules. Potential cooperation partners may coexist for months, years, decades, generations until they eventually chose to cooperate, if they ever do. The unpredictable and random development of real-world heterospecific cooperation cannot be reconstructed in an artificial laboratory setting, nor could a simulation provide information on whether cooperation between humans and wild crows has in fact been established somewhere at some point in the past, and if so, to what extent. However, the cognitive and acoustic prerequisites enabling this development can well be tested. The aim of the following experiment is therefore to examine whether these perceptual conditions are met by testing whether humans can recognize two individual large-billed crows by their ka-calls in a force-choice paradigm (without the option to defect due to the reason outlined above) and then link their identity to their behaviour (reciprocal cooperator or cheater). Based on the discrimination success in experiment 3 (unequal to recognition, but showing sensitivity to individual differences) and the fitness benefits associated with this ability in combination with cooperation, participants are expected to learn which calls belong to which crow, and then which crow belongs to which behavioural pattern. Note that this does not simulate a natural cooperation situation because participants do not have the option to defect.

2. Experiment 4: Individual Recognition in a Reciprocal Exchange Paradigm

2.1. Material and Methods

Subjects

Adults (N=18, 14 female). Sample size was initially set at 40 based on an a-priori power analysis using GPower version 3.1 (Erdfelder et al., 1996) with an estimated effect size dz of 0.45-0.5 for a minimum power of 0.8 with a two-tailed, paired-sample Wilcoxon signed-rank test (used to analyse cooperation decisions, see analysis). However due to the increasing spread of Covid-19, data collection was terminated on March 13th 2020 as part of the social distancing measures. As the resulting sample size was relatively small and high individual variation to be expected, the data collected up to this point was analysed as a case study where appropriate. Subjects were students and staff at Middlesex University between the ages 18 to 47 (mean age 22.8, SD=6.35), did not report hearing problems and gave informed consent. Approval of the ethics application had been obtained prior to data collection (see Appendix B, figure 3). Four participants had already participated in one of the previous discrimination experiments. First-year undergraduate Psychology students received credit points for their participation (N=7).

Stimuli

Stimuli were 16 set pairs of the ka-calls used in Experiment 3 (8 from each crow) matched for pitch (see Appendix A, table 9). Each pair consisted of one call from each crow, and the order of crows within the pairs was counterbalanced.

Apparatus

Identical to the previous experiments in Chapter 1.

Procedure

The experiment consisted of 30 rounds and pairs were chosen at random each trial to avoid predictability. The number of rounds is lower than in the previous three experiments due to the observed decline in performance towards the end of the experiment attributed to fatigue, but not as low as 20 rounds because this task was expected to be more difficult than the discrimination task. Before the experiment, participants received the following instructions: “Imagine you have some peanuts. You can exchange them for a more valuable object, a piece of bread. Your exchange partners are two crows. Each round, you hear both crows once. Then you have to decide which crow you want to give a peanut to, the first crow (1) or the second crow (2). The crow you choose can either give you the bread, or it can give you nothing. Your task is to get as many pieces of bread in the exchange as possible. After your decision, you will see which crow would have cooperated, the first (1) or the second (2)". They were not told that one of the crows would always reciprocate and the other would always defect. Their task was thus to recognize the two individuals by a single call and predict their behaviour in the exchange based on their experience from previous trials. As defecting was not an option, this is a forced-choice design. After participants made their decision (through either keypress "1" or keypress "2"), the computer displayed which crow would have participated in the exchange and reciprocated with a piece of bread, providing feedback to the participants about the correctness of their recognition. The ka-calls were the only way for participants to identify the crow and predict their behaviour. After the last round, participants were asked which strategy they used ("I chose the friendly crow", "I recognized voices", "I always picked the first crow", "I always picked the second crow", "I chose at random", "other").

Analysis

Cooperation decisions were recorded in PsychoPy3 (Peirce et al., 2019). The first 5 rounds were considered training rounds, as participants cannot possibly recognize a crow's call in the very beginning and correct decisions therefore depend on previous learning. The number of times participants cooperated with each crow after the training phase (rounds six to 30) was analysed with a paired sample, two-tailed t-test. Sex and musical background were not analysed due to the non-results in experiment 3 and the low sample size. Since "incorrect" strategies may indirectly lead to correct decisions (e.g. participant reports strategy based on perceived friendliness of the crow but consistently perceives crow *A* to be friendlier, thus recognized crow *A* without being aware of it), this variable was not statistically analysed and only served as a qualitative insight into the decision-making process. A potential learning trend throughout the experiment (including the training phase) was analysed with a linear regression model to see whether participants learned to only cooperate with the reciprocal crow.

2.2. Results

On average, participants chose the reciprocating crow (crow *A*) 11.38 times (SD=4.81, 95% CI [9.16, 13.6]), while they chose the defecting crow (crow *B*) 13.61 times (SD=4.81, 95% CI [11.38, 15.82]). Cooperation frequency with crow *A* ranged from 0% to 88% between participants (see figure 28). A paired sample, two-tailed t-test showed no significant differences between the absolute number of cooperation events with crows *A* and *B* (t = -0.97).

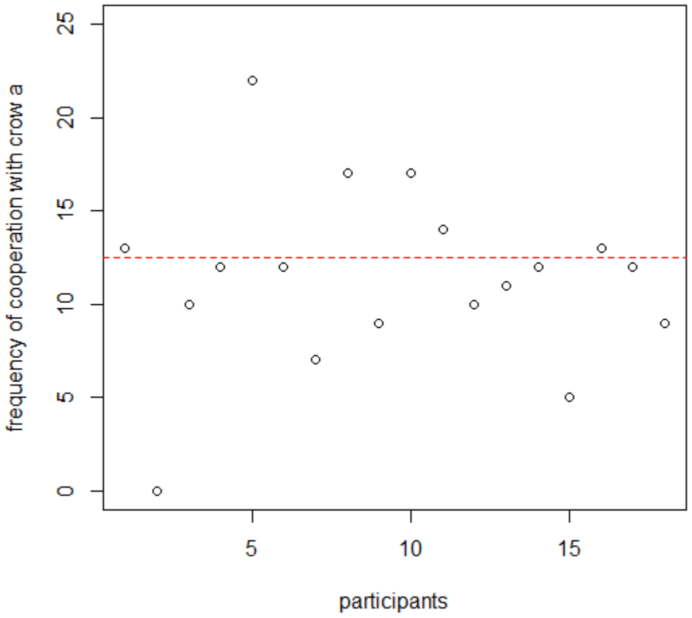


Figure 28: Frequency of cooperation events with crow *A* per participant in absolute numbers, created in R version 3.6.1 (R Core Team, 2019). Each point represents the total count of one participant. The horizontal line indicates chance level at 12.5 events.

Strategies employed by participants varied greatly inter-individually. Four participants reported choosing based on the crows' voices, seven chose based on perceived friendliness, three chose randomly, and four chose a strategy not prompted in the questionnaire (see figure 29). Out of the four participants who reportedly based their decision on the voice of the reciprocating crow, one did not choose crow *A* once across all trials, one chose both individuals equally, and the remaining two showed a tendency towards crow *A* and crow *B* respectively. Perceived friendliness of the crows was most often used and cooperation with crow *A* following this strategy ranged from five to 22 events across all trials. The participants following the random strategy, or another strategy not suggested in the questionnaire all performed around or below chance level. During the debriefing conversation, two participants who performed at chance level verbally reported (without being asked) to have recognized the "right" crow as the one with the shorter call duration. A third participant detailed to have followed the "tit for tat" strategy mentioned in the Introduction.

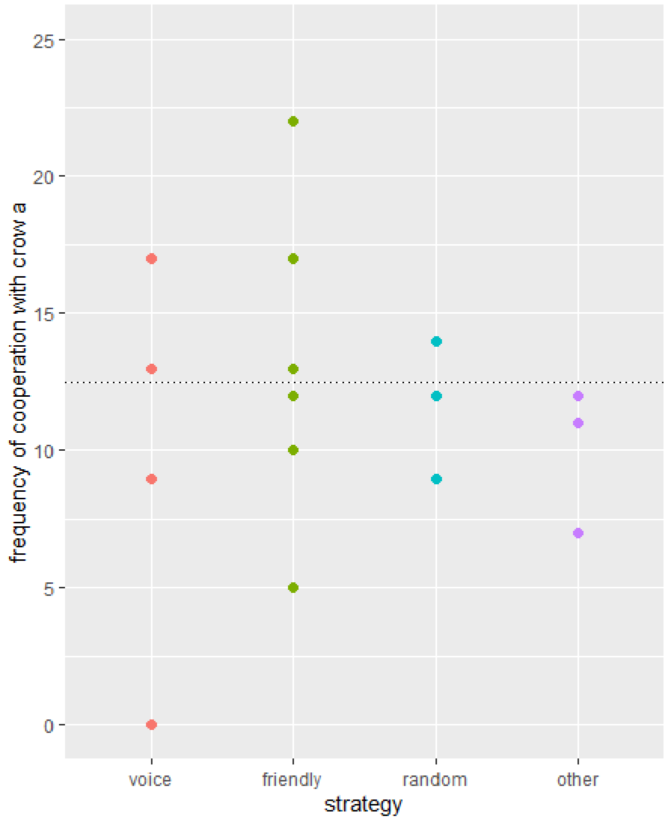


Figure 29: Frequency of cooperation with crow *A* in absolute numbers plotted per strategy (voice recognition, perceived friendliness, random, and other strategy) with one data point per participant, created in R version 3.6.1 (R Core Team, 2019) with the package ggplot2 (Wickham, 2009). The dotted line indicates chance level at 12.5 cooperation events.

A linear regression showed no significant, although positive trend across all trials (m=0.4; see figure 30).

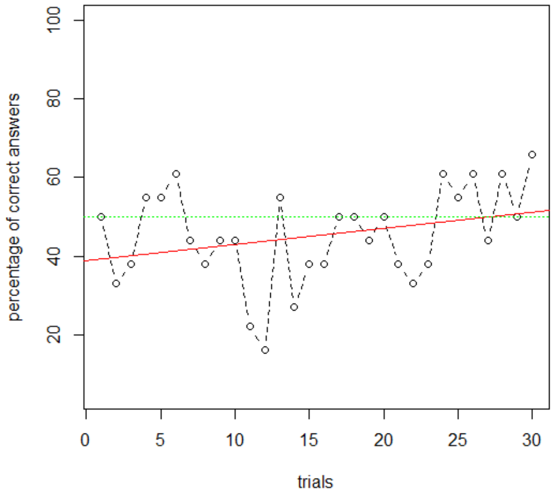


Figure 30: Percentage of correct answers (crow *A* chosen) for each trial with one answer per participant and trial, created in R version 3.6.1 (R Core Team, 2019). The horizontal line at 50% indicates chance level (green), and the linear regression line (red) indicates the overall trend.

3. Discussion

These results show that humans do not recognize individual crows by their ka-calls under the present conditions. The clear difference in performance in this experiment compared to the three previous experiments may be due to four factors: The first factor is that discrimination is an easier task as it only requires short-term memory of the two stimuli to be compared, followed by an assessment of similarity that is either above or below a threshold indicating separate origin. Recognition requires identification of relevant, stereotypical acoustic cues for each stimulus category, creation of a reliable mental template of these identifying cues, and retrieval of the templates for the categorization of subsequently presented stimuli. It may well be that discrimination is still easy enough to do without a priori knowledge or training, while recognition is not. This leads to the second factor, exposure. Recognition by definition is a learning-based ability and instead of directly comparing two stimuli, participants have to form a template of both voices based on single calls. It is therefore possible that participants would need more than these 30 trials to be able to recognize the two crows instead of simply discriminating them. Increased difficulty may also contribute to the third factor, motivation. The discrimination tasks were comparatively easy, although potentially boring. The increased difficulty in the recognition task may have led to frustration and thus a drop in motivation, especially in participants who were only moderately motivated for the task to begin with. The fourth factor is the comprehensiveness of instructions. Participants in the discrimination experiments received instructions detailing the entire extent of their task, and the task itself was presumably straight-forward. In this experiment, participants were not informed that the main focus would be on the vocal identity of the crows but were instead told to focus on the cooperation aspect of the task. While most participants will likely have guessed that the vocalizations should played some role in their choice, the aim of this study seems to have been clear only to the four participants following the voice strategy. Recognition accuracy may have been higher if the task had been as straight-forward as the discrimination task, without the cooperation decision to consider. Removing this part would of course also remove the ecological context and possible application of the results but would provide a more accurate insight into the pure recognition ability itself.

This drop in performance between the discrimination task and the recognition task provides insights into the limitations of our perception and processing of their calls. An example for this is the two participants who reported to have recognized the reciprocating crow by its relatively short call. All stimuli had the same length, and the mean duration of calls differed by 0.0016s between the two crows (see table 3). It would have been impossible for participants to discriminate, let alone recognize the crows based on the total duration of their calls. It is more likely that the participant perceived a different cue, such as a correlation between pitch and duration of call section but did not recognize it as such. The performance of the participant who consistently chose crow *B* in all 30 trials (including training trials) is particularly puzzling. It is highly unlikely that the participant chose the “wrong” crow by accident each time. The most probable explanation would be that they consistently thought the perceived nice crow would finally be the right choice despite contradictory evidence from the feedback from previous rounds. Even though it was the wrong crow in this task, the 100% accuracy in choosing the same crow in each round suggests that there is at least some stereotypical acoustic cue (or multiple cues) that could potentially be used in a simpler recognition task.

The differences in strategies further highlight the individual differences in each participant’s approach to these experiments. Differences in knowledge of game theory (such as the “tit for tat” strategy) or a participant’s familiarity with and opinion of the vocalizing animal may lead to very different outcomes. As none of the participants following the friendliness strategy chose the same crow in every trial, perceived friendliness may vary on a case-to-case basis and likely only supports the recognition. As already discussed in Chapter 1, it appears that many people find the notion of voice identity in a bird to be counterintuitive, which might explain why only a few participants in this study saw the connection between the playback of each crow’s call and its behaviour. Perceived friendliness is a more intuitive pattern for participants without background knowledge in this research area. As this is an entirely subjective category, strong deviations between the chosen crows are to be expected. Three participants following the friendliness strategy and two following the voice strategy show a tendency towards one crow over the other. While only three of them lean towards the reciprocating crow *A*, consistently choosing crow *B* also requires the recognition of some stereotypical cues in the stimuli (e.g. an acoustic cue that a participant perceives as communicating friendliness that is consistently present in the calls of crow *B*). Insights from this analysis of the different strategies employed are inherently limited and more data would be needed to statistically assess whether these tendencies are consistent or random pattern. In addition to that, the random strategy does not necessarily mean that a participant did not try to complete the task properly, it may also mean that they thought the crows’ behaviour itself was random. Strategies may also be combinations of two or more approaches, or it may be unclear to participants how exactly they chose. Future studies should also explore the potential species bias and use a more diverse range of animals, including human voices.

Based on these results it seems unlikely that humans would be motivated or able to make cooperation decisions with crows based on their vocal identity. Recognition in this experiment only requires memorizing for a short period of time, cooperation based on vocal identity in real life would require memorization and reliable recognition across a longer time span of weeks, months, or years. Since the discrimination of the vocalizations of this species has been successful and voice signatures have been shown to exist in this species (Kondo, Izawa, & Watanabe, 2010), the results are likely not due to species-specific reasons (such as obscure voice cues) but rather extend to other heterospecific vocalizations as well. As the sensitivity for differences in crow vocalizations (as assessed in Chapter 1) does not appear to be applicable to cooperation situations due to lack of ability, it remains unclear if and what advantage this sensitivity may have interspecifically. Kin recognition, pair bonds, parental investment, mate choice, and parental care are only relevant between conspecifics, leaving territoriality and co-habitancy with heterospecifics as a final possibility. Recognition of a neighbour would presumably require the same perceptual skill level as recognition of a reciprocating cooperation partner, which makes this possibility unlikely as well. Perceived friendliness may be useful to assess the threat level of an unknown heterospecific, but this would not rely on individuality. It currently appears most likely that the discrimination ability observed in Chapter 1 is based on a general perceptual ability that may primarily be used for conspecific discrimination and recognition, but that may be extended to heterospecific vocalizations, to a limited degree. It should therefore now be explored whether the extension of this potential general perceptual ability to heterospecific vocalizations is enabled by a shared origin of the relevant acoustic cues, or by the perceptual ability itself.

Chapter 3: Phylogeny

1. Introduction

1.1. Reconstructions of Stem-Amniote Vocalizations – Requirements and Limitations

The previous two chapters have illustrated that humans can easily discriminate two individual heterospecifics by their vocalizations, but not recognize them. I have hypothesized that this is enabled by a at least partially genetically determined, general perceptual ability (or multiple abilities) that can be applied to this task. This hypothesis requires one out of two conditions: Either amniote vocalizations contain a shared, homologous foundation of basic acoustic cues that is decoded by this general perceptual ability, or the perceptual ability is sufficiently developed to decode unfamiliar acoustic cues. Chapter 3 now focuses on the phylogenetic origin and subsequent development of amniote vocalizations to further explore the possibility of homologous voice cues. Fitch (2018) states that the "examination of patterns of similarity and difference relative to [a] phylogenetic tree allows us to distinguish examples of homology, in which a trait is shared by multiple species due to inheritance from a common ancestor, from analogy". He goes on that homology refers to the shared possession of a trait based on the occurrence of the trait in a common ancestor and the subsequent inheritance of the trait from said common ancestor, while analogy refers to the independent development of a similar trait across different species. The more species exhibit this shared trait, the earlier in time it likely emerged. Different concepts of homology can be applied to different features, all of which should be thought of as a scale of similarity rather than an absolute all-or-nothing categorization (partial homology). Homologous traits can be based on shared ancestry (plesiomorphy), classify a species or a group of species (apomorphy), be shared by the entire clade and indicate the inheritance of the trait from the last common ancestor or LCA (synapomorphy), or be specific of a species or group of species (autapomorphy). For instance, the larynx is considered a tetrapod synapomorphy due to the shared morphological structure across members of the clade, and the mammalian ability to reposition the larynx inside the vocal tract is a plesiomorphy. The descended larynx was initially thought to be a human autapomorphy (Fitch, 2018), but it has since been discovered in deer (Fitch & Reby, 2001), koalas (Charlton et al., 2011), and members of the subfamily Pantherinae (Weissengruber, Forstenpointner, Peters, Kübber-Heiss, & Fitch, 2002). As the source-filter theory appears to apply to all amniotes (Reber et al., 2015), sound production in the larynx (or syrinx), subsequent filtering in the vocal tract and the resulting voice cues could also be plesiomorphies. Following the results obtained in experiments 1 and 2, I will mostly focus on the acoustic cue pitch contour, which is made up of repetitive sequences of variations in fundamental frequency. For the pitch contour voice cue to be potentially produced by the stem-amniotes and thus potentially being an amniote plesiomorphy, the vocal tract of the last common ancestor of amniotes would have to fulfil at least the following three requirements: Vocal folds or a vibratory tissue comparable to the vocal folds, such as the labia within the syrinx (1; Riede & Goller, 2010a), individual differences in the size of the vocal folds (2), and sufficient control over the vocal folds to produce at least two different fundamental frequencies (resulting in high-low or low-high sequences of the individual's voice pitch; 3). Vocal fold control can be further divided into respiratory control of exhaled air (3a), and laryngeal muscle control (3b; Hirano, Ohala, & Vennard, 1969; Titze, 1994). As vocal fold size is correlated with overall body size (Riede & Brown, 2013) the second requirement can be considered fulfilled. If the remaining anatomical requirements are also fulfilled in the amniote LCA, a basic pitch contour system is conceivable. It should be noted however that this approach focuses exclusively on the production of voice cues, not the stem-amniotes ability to discriminate individuals or remember each other's voices, as that would exceed the insights gained from fossil reconstructions and comparative anatomical studies.

This approach also has a number of limitations. While fossil records provide an important baseline for the reconstruction of the LCA's vocal abilities, they remain largely incomplete. It is unclear how many remains of pre-historic species have been found, how many can possibly be found, and how well these remains reflect the species' actual morphological and physiological traits (Raup, 1972). Vocalizations do not fossilize at all and there is no certain way of knowing whether the last common ancestor of reptiles, birds, and mammals did in fact vocalize and if so, what information these vocalizations contained. As the larynx is made of cartilage (see Chapter 1, Vocalization - Production), which is softer than bone, it usually doesn't fossilise unless it is calcified (Bailleul, Hall, & Horner, 2012; Delpey, 1942) and there are currently "no fossils of the laryngeal sound source in frogs, reptiles or mammals" (Clarke et al., 2016). The absence of a larynx in the fossil record of an extinct pre-historic species is therefore no reliable indicator of the absence of a larynx in the living individuals. In addition to the uncertainty regarding the completeness of this reconstruction, the detection of general commonalities in vocalizations is further complicated by the similarity of vocalizations based on geographical closeness between two individuals, such as in the form of dialects in songbirds that show different levels of similarity between individuals of the same species (Nottebohm, 1975). Another factor influencing the reconstruction of vocalization similarities is the vocal adaption to the physical properties of the habitat. In areas with dense vegetation, lower frequencies are less scattered by the leaves and branches (Chapuis, 1972), and habitat conditions like these select for acoustic properties in vocalizations that may not reflect the state of acoustic properties in the vocalizations of an ancestor that lived in a different habitat. Vocalizations of vocal learner species undergo additional changes through cultural evolution (Payne, 1981). Other features could be atavistic and appear to be absent in modern species, even though they were present in an ancestor but are no longer expressed in the phenotype and thus not observable (Tomić & Meyer-Rochow, 2011). Reconstructions of the stem-amniotes vocal abilities through comparative analysis of modern amniotes are further limited as these species are only a subset of the descendants of the LCA. Many terrestrial species, including all non-avian dinosaurs (Fastovsky & Sheehan, 2005), 93% of mammalian species (Longrich, Scriberas, & Wills, 2016), 83% of lizards and snakes (Longrich, Bhullar, & Gauthier, 2012), and most archaic (not belonging to the clade Neornithes) bird species (Longrich, Tokaryk, & Field, 2011) became extinct during the Cretaceous-Paleogene mass extinction roughly 65.5mya due to the impact of the Chicxulub asteroid and the resulting environmental changes (Schulte et al., 2010), potentially resulting in a bottleneck effect on vocalization diversity. Due to these inherent uncertainties in the reconstruction of voice cue evolution, there can be no definitive conclusions on the presence or complexity of voice cues in the vocalizations of the LCA. Or as Fitch (2009, p.115) phrased it, “After careful consideration of various potential fossil evidence to speech, I will conclude that the best of the proposed cues provide imperfect clues […], and that most proposed cues tell us nothing”. The following reasoning is therefore inherently incomplete and, while based on the current state of knowledge founded in fossil records and comparative analyses, merely a collection of hypotheses. Dating of time periods is based on the International Chronostratigraphic Chart (K. M. Cohen, Finney, Gibbard, & Fan, 2013).

1.2. Stem-Amniotes

The first terrestrial vertebrates emerged in the Late Devonian (Carroll, 1992) and subsequently split into amphibians and anthracosaurs. These stem-tetrapods are thought to have possessed a simple larynx (Fitch 2018). Its main function was likely as a valve for the airway, although it already had a role in sound production in the subsequent Lissamphibia (Fitch 2016). Descendants of anthracosaurs, the stem-amniotes, were the first to reproduce outside the water, either through the form of laid eggs or internally developing embryos, distinguishing them from amphibians (Zug et al., 2007). The stem-amniotes diverged into the Synapsidia (today’s mammals) and the Reptilia 331–319 mya during the late Carboniferous (Ford & Benson, 2020). The Carboniferous was a time period characterised by lycophyte forests consisting of trees that reached a height of up to circa 45m and diameters of 2m (Thomas & Cleal, 2018), which likely resulted in compromised visual contact between individuals. In this habitat, acoustic communication would be beneficial for the stem-amniotes to contact conspecifics, at the very least in the form of simple mating calls.

Jarvis (2019) suggests that “the common ancestor of vertebrates had a brainstem pathway for [the] production of innate vocalizations with limited vocal plasticity, such as the Lombard effect, where animals increase sound production volume or pitch in noisy environments". Fitch (2018) specifies that while the basic anatomy of the larynx is shared between tetrapods, "the mammalian larynx is unique in possessing a thyroid cartilage and a thyroarytenoid muscle contained within the vocal folds". Movement of the thyroid cartilage relative to its distance to the cricothyroid muscle influences the voice pitch, but is less important for pitch than vertical laryngeal movement (Hong, Hong, Jun, & Hwang, 2015). The muscles connecting the hyoid, the laryngeal cartilages and the valve tissue (vocal folds when present) are largely homologous in tetrapods (Negus 1949, as cited in Kingsley et al., 2018). These muscles and the brainstem pathway enabling innate vocalizations and potential pitch modulations suggest that the third condition, movement of the vocal folds, may have also been fulfilled in the stem-amniotes, at least in a very basic form. Although the vocal anatomy of the stem-amniotes was far less complex than that of modern amniotes and research on the evolution of speech often contemplates significant physiological modifications as a prerequisite for speech, Fitch (2017) states that "there is little evidence that any major changes in the vocal apparatus itself were required for our ancestors to gain the capacity for speech" and instead considers "changes in complex vocal control rather than vocal anatomy [to be] the key innovations required for the evolution of speech". Macaque monkeys for instance have been found to possess a "speech-ready vocal tract" while still unable to speak, and so instead of physiological changes of the vocal tract, the evolution of speech relied on neural changes leading to a "speech-ready brain" (Fitch, Boer, Mathur, & Ghazanfar, 2016). In addition to the potential production of vocalizations differing in fundamental frequency, stem-amniotes likely possessed small, simple ears that included "hair cells with a cochlear amplifier mechanism, electrical frequency tuning, and incipient micromechanical tuning" (Manley, 2000). The cochlear amplifier mechanism amplifies sound vibrations and thus improves hearing (Oghalai, 2004), while electrical frequency and micromechanical tuning maximise the cell's response to different frequencies (Allen & Neely, 1992; Hudspeth & Lewis, 1988). Taken together, this shows that stem-amniote vocalizations with minimal pitch variance can be considered a realistic possibility, but it is unclear whether these vocalizations were sophisticated enough to contain pitch contour that allows for individual discrimination or recognition. To explore the possible levels of complexity, I will reconstruct the vocalizations of ancestral reptilian clades as much as possible based on the vocalizations of modern birds (Neoaves) and their phylogenetic relationships, starting with the focal species of this thesis, the zebra finches and large-billed crows.

2. Phylogenetic Reconstruction of Pitch Contour

2.1. Pitch Contour in the Neoaves Clade

Zebra finches (*Taeniopygia guttata*) belong to the family Estrildidae (Swaddle, 2010), which emerged approximately 20 mya. The genera closest to the *Taeniopygia* are the *Poephila*, *Emblema* and *Neochima*, which diverged from their last common ancestor 10.5 mya (Arnaiz-Villena et al., 2009). Members of the genus *Poephila* vocalize to identify and remain in contact with conspecifics, and so their songs vary on an inter-individual level. In addition to their song, 12 call types consisting of complex notes with harmonics have been classified (Zann, 1976). Species within the genera *Emblema* and *Neochima* are very social and produce distance calls to stay in contact with their flocks (Immelmann 1982 as cited in Forshaw & Shephard, 2012). The distance calls of the red-browed firetails (*Neochima temporalis*) closely resemble the distance calls of superb blue wrens (*Malurus cyaneus*) and individuals may also respond to heterospecific calls (Rowley 1965, as cited in Forshaw & Shephard, 2012). Their simple song consists of five notes and varies inter-individually (Goodwin 1982, as cited in Forshaw & Shephard, 2012). All of these species have been found to forage in heterospecific groups (Forshaw & Shephard, 2012).

The divergence of this group of genera and the ancestor of the remaining estrildid species (with the exception of the genera *Vidua*) is estimated to have happened 16.5 mya (Arnaiz-Villena et al., 2009). The songs of the genera within the family Estrildidae generally serve a contact function and a display function (Harrison, Nicolai, Immelmann, & Wolters, 1962). Song complexity within the family varies and ranges from relatively simple two-note songs in black-cheeked waxbills (*Estrilda erythronotos;* Immelmann et al., 1965, 1977 as cited in Baptista & Trail, 1992) to more complex songs consisting of multiple harmonics in Gouldian finches (*Chloebia gouldiae*, Thorpe 1961 as cited in Baptista & Trail, 1992). The genus *Lonchura* is particularly of interest for comparative studies on language evolution. The Java sparrow (*Lonchura oryzivora*, but also classified as *Padda oryzivora*) has been found to discriminate English and Chinese sentences spoken by a bilingual speaker (Watanabe, Yamamoto, & Uozumi, 2006), and is able to discriminate prosodic patterns in speech (Naoi, Watanabe, Maekawa, & Hibiya, 2012), as are zebra finches (Spierings & ten Cate, 2014). Bengalese finches (*Lonchura striata var. domestica*) can use prosodic cues to segment songs into note-chunks (Takahasi, Yamada, & Okanoya, 2010). The quail finch (*Ortygospiza atricollis*) produces a variety of different call types, including contact calls that can consist of a single note or a series of notes, and male vocalizations include variations in pitch contour. Both male and female vocalizations vary inter-individually (Nuttall, 1993). Strawberry finches (*Amandava*

amandava) produce songs that consists of multiple syllables of descending pitch, which is followed by a whistle and a trill (Baptista & Petrinovich, 1984).

Large-billed crows belong to the genus *Corvus* within the family Corvidae and the superfamily Corvoidea, which originated roughly 33.5 to 33.8 mya (Ericson, Jansén, Johansson, & Ekman, 2005). Based on a Bayesian analysis of molecular sequences, Jønsson, Fabre, and Irestedt (2012) suggest that the genus *Corvus* originated roughly 17.5 mya in the Palaearctic region from where they radiated to most continents. Large-billed crows likely emerged 2.6 to 5.3 mya (Haring, Däubl, Pinsker, Kryukov, & Gamauf, 2012). They are most closely related to Mariana crows (*Corvus kubaryi;* Jønsson et al., 2012). Tomback (1986) observed four types of vocalization in this species: High pitched caws used by all flock members to locate each other, nasal caws used only by mated pairs to locate each other, monologues produced by pair members as courtship or pair-bond maintenance, and sharp alarm call caws. If complete, this call repertoire is considerably smaller than that of other crows who also produce scolding calls, dispersal calls (Chamberlain & Cornwell, 1971), mobbing calls (Yorzinski & Vehrencamp, 2009), sequential note-calls establishing hierarchy relations (Kondo & Hiraiwa-Hasegawa, 2015), or calls advertising a food source (Soma & Hasegawa, 2003). Pitch contour differences contribute to the variation between call types produced in different contexts in American crows and individuals can be identified based on their calls (Mates, Tarter, Ha, Clark, & McGowan, 2015). This is similar to the cawing format of the carrion crow (Thompson, 1982), and the signature voice system in large-billed crows in which differences in pitch and temporal cues in ka-calls indicate identity (Kondo, Izawa, & Watanabe, 2010). The sensitivity to voice cues in carrion crows extends to the discrimination of familiar and unfamiliar human voices and jackdaw (*Corvus monedula*) calls (Wascher et al., 2012). Like the crows, common ravens produce a variety of different call types that differ in their fundamental frequency depending on the vocalization context (Conner, 1985). Outside the genus *Corvus* but within the family Corvidae, the pinyon jays (*Gymnorhinus cyanocephalus*) produce different types of vocalizations that differ in their pitch contour and allow for inter-individual recognition, such as the differences in pitch contour in the near call (Berger & Ligon, 1977).

This comparative review shows that complex vocalizations with pitch contour as at least one of the discriminating parameters are commonly found in the families of zebra finches and large-billed crows. Both the Estrildidae and the Corvidae belong to the suborder Passeri (oscines) within the order Passeriformes. This order emerged roughly 47 mya and also includes the suborders Acanthisittidae and Tyranni (suboscines). Tyranni and Passeri diverged roughly 44 mya, and the crown Passeri emerged roughly 38 mya (Oliveros et al., 2019). Vocal learning is not thought to be present in these two suborders and was either lost in both or emerged separately in the Passeri (Jarvis et al., 2014). Without vocal learning, the species within the

Tyrannidae and the Acanthisittidae rely on innate vocalizations (Päckert, 2018). The songs of tyrant flycatchers (Tyranni), such as the Acadian flycatchers (*Empidonax virescens*), show little inter-individual variation, and there appears to be no auditory recognition between neighbours (Wiley, 2005). When faced with anthropogenic noise, which prompts other birds to alter the frequency of their song (Slabbekoorn & den Boer-Visser, 2006), grey flycatchers (*Empidonax wrightii*) reduce their occupancy of the area instead of altering their songs. Ash-throated flycatchers (*Myiarchus cinerascens*) minimally increased their fundamental frequency, although that may simply be a secondary effect of the increased amplitude (Francis, Ortega, & Cruz, 2011). The two species within the Acanthisittidae, the New Zealand rock wren (*Xenicus gilviventris*) and the rifleman (*Acanthisitta chloris*) each produce two different call types containing ultrasonic (>20 kHz) harmonics, while the highest fundamental frequency recorded was 12.4 kHz and 11.3 kHz respectively (Krull, C. R., Parsons, S., & Hauber, M. E., 2009; Lloyd-Jones, 2014). The spectrograms of the calls of the male and female New Zealand rock wrens provided show very similar pitch contours. In the cooperatively breeding rifleman, contact calls are significantly more similar between relatives, but two different playback experiments of kin and non-kin did not elicit an aggressive or affiliative response from individuals (Hodges, 2012; Khwaja, Briskie, & Hatchwell, 2019).

This apparent absence of individual recognition and discernible voice cues in the suborders Tyranni and Acanthisittidae raises the question whether they are generally absent in Neoaves and only present in vocal learners. Following the phylogenetic tree of Neoaves proposed by Jarvis et al. (2014) based on a whole-genome phylogenetic analysis, the Neoaves consist of the three vocal learner orders and sub-orders Passeri, Psittaciformes (parrots) and Caprimulgiformes (hummingbirds), as well as 35 non-vocal learner orders and sub-orders. The authors suggest that vocal learning emerged separately in two or three lineages, either separately in parrots, hummingbirds and oscines, or in last common ancestor of oscines and parrots, separately in hummingbirds, and followed by a secondary loss in New Zealand wrens and suboscines (who share a common ancestor with oscines and parrots, see figure 31). Out of the groups included in this tree, 14 have been found to produce vocalizations that allow for individual discrimination or recognition based on pitch contour or a set of frequency parameters comparable to pitch contour. Four have been found unable to do so, and the remaining 20 could not be categorized due to lack of literature on this specific subject. As hummingbirds and parrots are vocal learners that acquire their song from other individuals, their songs also vary inter-individually (Berg, Delgado, Cortopassi, Beissinger, & Bradbury, 2012; Masin, Massa, & Bottoni, 2004; Yang, Lei, Wang, & Jesse, 2007). Within the Accipitridae (emerged ~ 60 mya), alarm calls of the white-tailed hawk (*Buteo albicaudatus*) differ between individuals, including differences in pitch contour (Farquhar, 1993). Individual bald eagles (*Haliaeetus leucocephalus*) can be discriminated based on pitch characteristics of their calls (Eakle, Mannan, & Grubb, 1989).

Individual European eagle owls (*Bubo bubo*, Strigiformes ~64 mya) can be identified using multiple acoustic parameters in their calls, including pitch contour (GRAVA, Mathevon, PLACE, & BALLUET, 2008) and tawny owls (*Strix aluco*) can discriminate neighbours from strangers (Galeotti & PAVAN, 1993). Penguins (Sphenisciformes, ~64 mya) are able to recognize individuals based on their vocalizations, presumably because they live in large colonies of visually hard to recognize individuals. Magellanic penguins (*Spheniscus magellanicus*) discriminate conspecific calls, females recognize the call of their mate, and chicks recognize the calls of their parents (Clark, Boersma, & Olmsted, 2006). King penguin (*Aptenodytes patagonicus*) chicks recognize their parents based on the pitch contour of their vocalizations (Jouventin et al., 1999). Red-legged seriemas (*Cariama cristata*) produce calls that differ inter-individually in their frequency and consist of notes with distinct pitch contour patterns (Padget, 2010). The provisioning calls of the European bee-eater (*Merops apiaster*) vary significantly in their pitch contour between individuals (Lessells, Rowe, & McGregor, 1995). Individual Rufous-headed Hornbills (*Aceros waldeni*) produce calls that differ in their frequency parameters (POLICHT, PETRŮ, LASTIMOZA, & SUAREZ, 2009). The territorial yodels of male common loons (*Gavia immer*) differ inter-individually in the frequencies of the highest intensity at specific points in the introduction (comparable to pitch contour, which was not tested for; Walcott, EVERS, FROEHLER, & KRAKAUER, 1999) and males can discriminate between the yodels of neighbours and strangers (Mager III, Walcott, & Piper, 2010). The long-distance advertisement calls of male piping plovers (*Charadrius melodus*) varies between individuals based on the pattern of the dominant frequency (Sung & Miller, 2007). The recognizability of calls in red-crowned cranes (*Grus japonensis*) fluctuates throughout their lives and is highest for chicks when families gather in flocks and offspring may get lost in the crowd, and pitch contour plays an important role in their recognition (Klenova, Volodin, & Volodina, 2009). Male North African houbara bustards (*Chlamydotis undulata undulata*), who rely on lekking courtship, can be recognized by their booms mostly based on the maximum frequency, as well as the frequency in relation to the harmonics and the energy. The slope of the mean frequency (comparable to pitch contour) showed slight individualisation (Cornec, Hingrat, & Rybak, 2014). Male common cuckoos (*Cuculus canorus*) can be distinguished by their advertisement calls based on pitch contour (Zsebők, Moskát, & Bán, 2017). Acorn woodpeckers (*Melanerpes formicivorus*) produce waka-calls consisting of syllables that can be used to recognize individuals by a hidden Markov models, but it was not investigated whether pitch contour contributes to this process (Yao, Lin, Ali, & Taylor, 2006). Davis (1980) describes some inter-individual variation in the calls of multiple heron species, but the data presented is not enough to make definitive conclusions. In a playback experiment, Northern bald ibises (*Geronticus eremita*) reacted slightly more often than chance level to the greeting croops of their mate and less often to the calls of non-mates indicating some individual recognition, but pitch contour was not identified as a relevant cue in this experiment (Szipl, Boeckle, Werner, & Kotrschal, 2014). The calls of male European nightjars (*Caprimulgus europaeus*) have been found to be unstable over time and thus do not provide a reliable basis for individual recognition (Raymond et al., 2019). While the greater flamingo (*Phoenicopterus ruber*) can be acoustically recognized, pitch contour is not relevant (Mathevon, 1996).

Taken together, pitch contour as a voice cue is found in all Neoaves clades except for the Columbea (the clade including the flamingos and doves; see figure 31). No literature on inter-individual pitch contour recognition in the other bird clades outside the Neoaves, the Galloanseres (landfowls and waterfowls) and the Palaeognathae (tinamous and ostrich) was found. It is unclear whether absence of literature on this subject is due to lack of interest in this trait in the species in question, difficulty of accessing these species, or because negative results simply haven't been published.

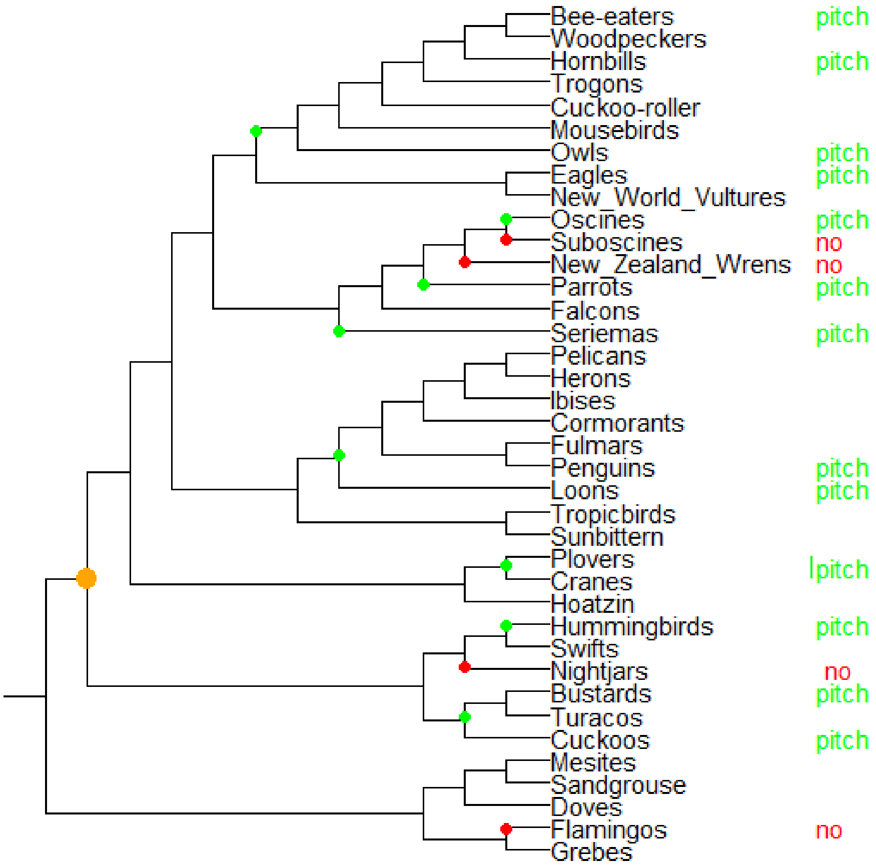


Figure 31: Phylogenetic tree of Neoaves based on (Jarvis et al., 2014) created in R version 3.6.1 (R Core Team, 2019) with the packages "ape" (Paradis E. & Schliep K., 2018) version 5.3 and ggtree (G. Yu, Smith, Zhu, Guan, & Lam, 2017). Green pitch labels on the right indicate that at least one species within the taxon has been found to produce pitch contours that can be used to discriminate or recognize individuals. Red labels indicate that pitch contour was either found to be absent or irrelevant. Taxons without labels have not been tested experimentally yet. The orange dot indicates the hypothesized minimum point of emergence for pitch contour, green dots indicate minimum presence of pitch contour, and red dots indicate minimum absence. Branch lengths are not scaled to years.

Since pitch contour as a voice cue is found frequently within the Neoaves and is evenly spread between the Passerea (all Neoaves clades except for the Columbea), it was likely already present in the last common ancestor of the Passerea, and secondarily lost in the nightjars, New Zealand wrens, and suboscines (rather than emerging a minimum number of 7 times in the various nodes, or up to 14 separate times in each species). It is currently unclear whether the pitch contour cue is absent in all Columbea, or whether it was secondarily lost in the flamingos and remains to be found in the other species. Pitch contour in the last common ancestor of all Neoaves can therefore be neither supported nor rejected at this point.

Prum et al. (2015) propose a different phylogenetic tree for Neoaves that would suggest a different pattern of pitch contour in the diverse families (see figure 32).

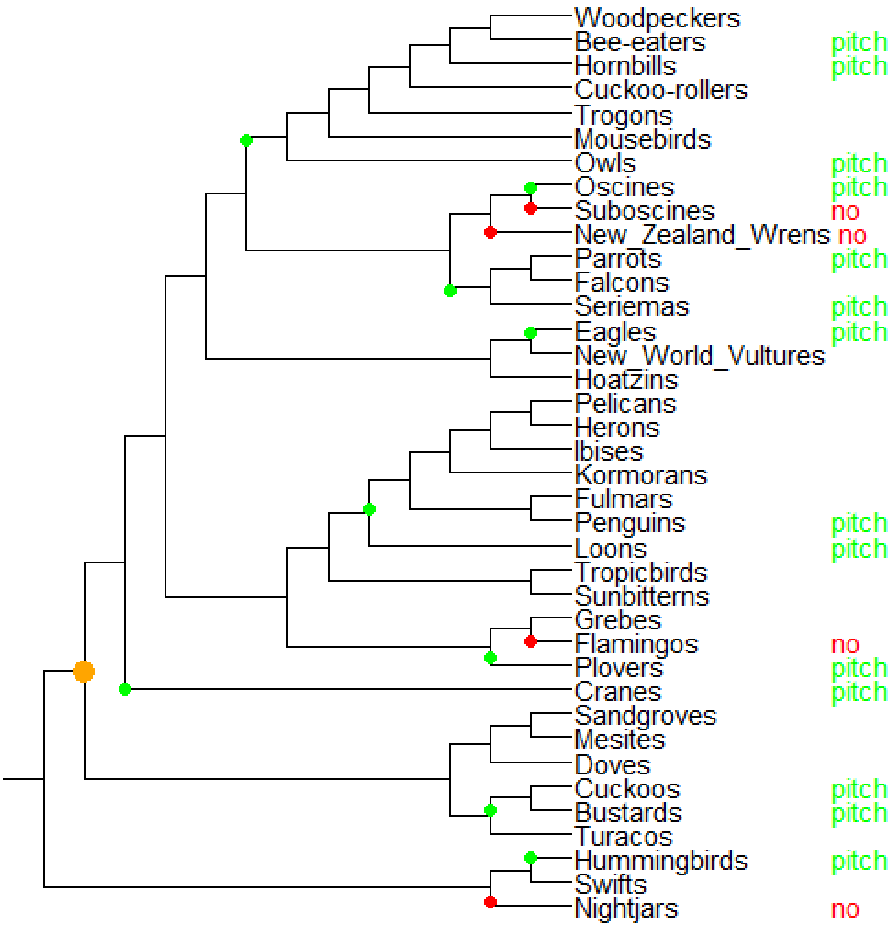


Figure 32: Phylogenetic tree of Neoaves based on (Prum et al., 2015) created in R version 3.6.1 (R Core Team, 2019) with the packages “ape” version 5.3 (Paradis E. & Schliep K., 2018) and ggtree (G. Yu et al., 2017). Green pitch labels on the right indicate that at least one species within the taxon has been found to produce pitch contours that can be used to discriminate or recognize individuals. Red labels indicate that pitch contour was either found to be absent or irrelevant. Taxons without labels have not been tested experimentally yet. The orange dot indicates the hypothesized minimum point of emergence for pitch contour, green dots indicate hypothesized minimum presence of pitch contour, and red dots indicate minimum absence. Branch lengths are not scaled to years.

This alternative tree suggests that the voice cue pitch contour either emerged in the last common ancestor of all clades except the Strisores (the clade including the nightjars), individually in the hummingbirds and was then secondarily lost in flamingos, New Zealand wrens and suboscines, or that it emerged in the last common ancestor of all Neoaves and was then secondarily lost in the nightjars, flamingos, New Zealand wrens, and suboscines. Solving this ambiguity would require an experiment on pitch contour as a voice cue for individuality in the swifts. In both trees, the trait appears to be present at least after the first divergence within the Neoaves. Prum et al. (2015) suggest that the Strisores diverged first within the Neoaves shortly before the K-Pg boundary, and Jarvis et al. (2014) propose the same for the Passerea. These are of course entirely different clades and it is unclear which phylogenetic analysis is more accurate, but both trees support the hypothesis that pitch contour is commonly found in modern Neoaves across diverse clades and is possibly a plesiomorphy of these clades, and that it was already present before the K-Pg boundary roughly 66 mya.

2.2. Pitch Contour in Archaic Birds and non-Avian Dinosaurs Before the K-Pg Boundary

Four archaic bird groups, namely Hesperornithes, Ichthyornithes, and Palintropiformes, and the Ornithurae sister clade Enantiornithes lived until the K-Pg boundary. Few Ornithurae members are known to have survived the mass extinction, including Ornithurine C (Longrich et al., 2011), *Maaqwi cascadensis* (McLachlan, Kaiser, & Longrich, 2017) and *Vegavis iaai* (Clarke, Tambussi, Noriega, Erickson, & Ketcham, 2005). Ornithurine C is closer to the Neornithes than to the Ichthyornithes, although it is currently unclear which of these two clade it actually belongs to (Longrich et al., 2011). *Vegavis iaai*, which belonged to the clade Vegaviidae (Agnolín, Egli, Chatterjee, Marsà, & Novas, 2017), is the only known member of the Neornithes to survive the K-Pg boundary (Marsà, Agnolín, & Novas, 2019) and is considered to be a basal Anseriformes clade. The Anseriformes do not belong to the Neoaves and diverged roughly 72 mya (Prum et al., 2015).

The oldest currently known fossilized syrinx consists of nine mineralized rings and has been traced back to the Late Cretaceous 66-69 mya, shortly before the K-Pg boundary. It belonged to *Vegavis iaai* and the positions of the rings allow inferences about the location of the sound-producing tissues. Its overall structure and the phylogenetic reconstruction of the species suggests that *Vegavis* could produce vocalizations similar to those produced by ducks or geese. The absence of fossilised syrinxes from other species in this and previous time periods, despite the apparent fossilisation potential of the syrinx, may suggest that it had not evolve until this point, although further evidence for this would be required (Clarke et al., 2016). In addition to the evolution of the novel syrinx, the avian larynx shifted in its position during this time period.

Li, Zhou, and Clarke (2018) outline a gradual laryngeal shift across the dinosaur phylogeny, and hypothesize that this trait may have been minimally present in the Paraves, the avian ancestor. In their review of the evolution of the syrinx, Kingsley et al. (2018) propose two possible explanations why modern birds do not vocalize with the larynx anymore: The vocalizing function of the larynx was lost in the stem-birds and then the syrinx evolved to replace this function (implying a quiet transitional period without vocalizations), or the syrinx evolved as an improvement to the still vocalizing larynx, which was subsequently replaced as the vocal organ by the syrinx. Riede et al. (2019) support the latter possibility by suggesting that the selective pressure driving the evolution of the syrinx was the increased efficiency of sound production achieved, such as multi-layered vocal folds in some avian species (Riede & Goller, 2014). The ancestral syrinxes likely produced low fundamental frequencies within a small frequency range due to limited vocal control (Riede et al., 2019). This initially low fundamental frequency presumably did not require signature vocalizations, but the following gradual increase in vocal control and fundamental frequency may have contributed to the development of signature songs vocalizations in later species, as voice cues are less reliable in high frequency vocalizations (Fitch, 1997).

The currently oldest known common ancestor of birds is *Aurornis xui*, an avialan (transitional between non-avian dinosaurs, called dinosaurs in the following, and true birds) species that emerged before *Archaeopteryx* and *Anchiornis* roughly 170 mya (Godefroit et al., 2013). The dinosaurs can be divided into two major groups, the Ornithischia and the Saurischia. The latter includes the theropod dinosaurs and thereby the birds, together with the *Tyrannosaurus* or the *Velociraptor* (Weishampel, Osmlska, & Dodson, 2004). The Lambeosaurinae emerged roughly 85 mya (Paul, 2010) and belonged to the group Ornithischia that lived 66 – 200 mya (Weishampel, 2004). Based on a resonance analysis of their crest containing parts of the nasal cavity, as well as a review of their auditory anatomy, Lambeosaurinae (Late Cretaceous, Hadrosauridae family) likely produced vocalizations with formants. Sexual dimorphism of the crest, as well as differences in frequency between adults and juveniles, and sensitivity to high-frequency vocalizations, suggest that these vocalizations were potentially relevant for parent-offspring communication and under sexual selection (Weishampel, 1981). Although the nasal resonance effect alone is no guarantee that the Lambeosaurinae did in fact vocalize, it is quite possible considering that the dinosaurs likely had a simple larynx (see Kingsley et al., 2018 and Clarke et al., 2016). Resonance effects have been considered as a possible function of the nasal cavity of other dinosaurs as well, such as anklyosaurs (Witmer & Ridgely, 2008). Based on their phylogenetic reconstruction of closed-mouth vocalizations in archosaurs, the group including crocodiles and birds, Riede and colleagues (2016) conclude that at least some dinosaurs likely produced both open-mouth and closed-mouth vocalizations. Following these reconstructions, the vocalizations of at least some Sauropsidia and the Ornithischia appear to have included the acoustic features fundamental frequency and formants. The authors do not suggest a time period for this trait, but since this reconstruction is based on the vocalizations of modern Archosaurs, the crocodiles and birds, a broad reference point to suggest may be the mid-Triassic 247-237 mya (K. M. Cohen et al., 2013) when the common ancestor of birds and crocodiles lived (Weishampel, 2004). From this data it is however unclear whether basal Archosaurs produced vocalizations with a monotone pitch, species-stereotyped pitch, or individually stable pitch contours. While research on individual recognition in crocodilians is limited, individual signatures in juvenile Nile crocodiles have been found to be too weak for individual recognition (Vergne, Avril, Martin, & Mathevon, 2007), and there is currently no evidence for individual pitch contour cues in ancestral Archosaurs clades.

3. The Emergence of Pitch Contour

Taken together, these reconstructions provide the following hypothetical evolutionary timeline for the emergence of pitch contour as a voice cue signalling identity (see figure 33).

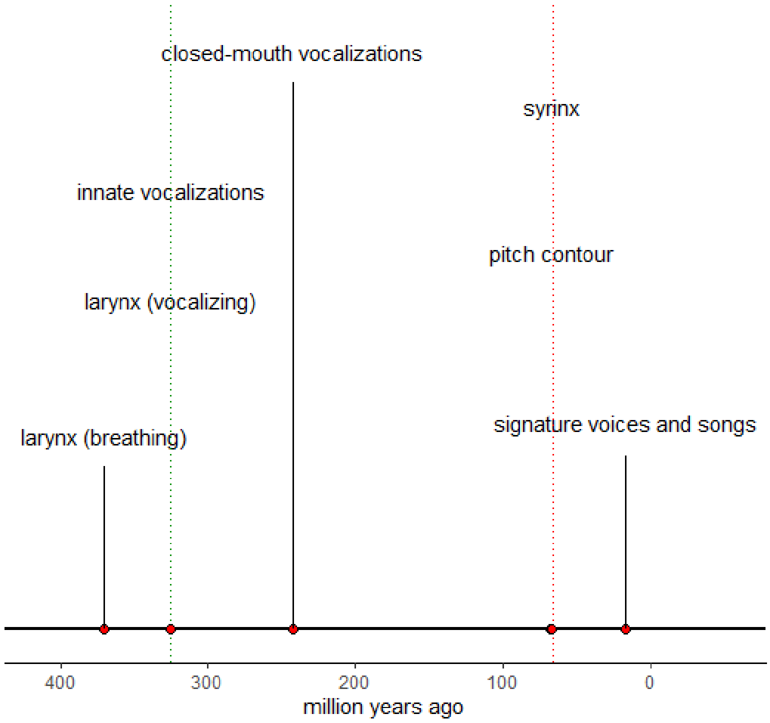


Figure 33: Proposed chronological timeline for the emergence of the most important traits related to the pitch contour voice cue discussed above, created in R version 3.6.1 (R Core Team, 2019) with the package ggplot2 (Wickham, 2009). The red line marks the mass extinction at the K-Pg boundary, the green line the divergence of the amniote LCA into the ancestors of Diapsids and Synapsids. The most recently emerged trait is the signature system employed by members of the Estrildidae and Corvus, preceded by the syrinx and the pitch contour in a basal Neoaves species shortly before the K-Pg boundary. The date for the emergence of closed-mouth vocalizations is the median of the time range proposed above. The earliest traits are the stem-amniote larynx that is thought to have produced innate vocalizations, preceded by the simpler stem-tetrapod larynx that likely could not vocalize yet.

The reconstructions discussed above suggest that the voice cue pitch contour was already present in a basal Neoaves species shortly before the K-Pg boundary, but possibly not in the last common ancestor of all Neoaves. The syrinx appears to have evolved roughly one million year prior to that, as it was present in a basal Neornithes clade ancestral to the Anseriformes that emerged before the Neoaves. The dinosauria likely produced closed-mouth vocalizations (possibly in addition to open-mouth vocalizations) with a larynx, a trait that may have emerged in the Archosaurs 247-237 mya. The amniote LCA is thought to have produced innate vocalizations with a simple larynx and some, although limited vocal plasticity. These time periods indicate the minimum age of a trait, some or all of them may date back further. Nevertheless, there is currently no evidence supporting the hypothesis that pitch contour as a voice cue could be an amniote plesiomorphy. Instead, it is likely an analogous trait that evolved separately in the Neoaves and the Synapsids but is acoustically similar and thus perceivable across species due in part to the similar morphology of the larynx and the syrinx, which both follow the principles outlined by the source-filter theory. Future studies may try to disentangle the contributions of these morphological similarities and the extent to which human perception tolerates exclusively heterospecific cues.

Conclusion

My primary aim was to examine the perception of individual identity in heterospecific vocalizations by humans. Chapter 1 first established the mechanism of amniote vocalizations as a proximate cause, discussing the similarities in their production and heterospecific perception of acoustic cues. Experiments 1 and 2 showed that humans can discriminate two individual zebra finches with high accuracy, and that this discrimination is largely based on differences in pitch contour, although this is not the only relevant cue. Experiment 3 showed that this is also possible with the calls of large-billed crows, although pitch contour could not be confirmed as a relevant discrimination cue. Discrimination accuracy was highly variable inter-individually, and neither sex nor musical background seemed to play a role in discrimination success. Accuracy was already high in the first trials but did not significantly improve with practice. I therefore hypothesized that discrimination was accomplished with a general perceptual ability that is at least partially genetically determined and co-opted for this task. Future studies should explore what other acoustic cues are involved in this discrimination, and why the discrimination of these two species was more successful than that of rhesus monkey voices as reported by (Friendly et al., 2014). Chapter 2 focused on the functions of this ability as an ultimate cause, both in the context of conspecific and heterospecific discrimination and recognition. The central focus was on reciprocal altruism between heterospecifics but experiment 4 did not provide any evidence that this could be a relevant function for the recognition of individual heterospecifics. More work is necessary to determine whether humans could recognize individual crows in a simpler task without cooperation decisions and with more training, as well as narrow in on the effects of perceived friendliness and other potential biases. Chapter 3 then discussed the possibility of homologous amniote voice cues by reviewing the phylogeny of birds and tracking the evolutionary origin of pitch contour in bird song as the second ultimate cause. Since pitch contour likely emerged separately in the Neoaves and the Synapsids, this voice cue appears to be analogous in amniotes. It should be noted that this reconstruction only applies to pitch contour, as the other relevant voice cues could not be identified here, and that the accuracy of phylogenetic reconstructions are inherently limited by the incomplete literature. As such, only minimum emergence but not maximum emergence time can be estimated at this time. Discrimination might be possible despite this separate origin due to similar morphology of the vocal organs that both follow the principles outlined in the source-filter theory.

Future studies may want to identify the perceptual ability involved in the discrimination and recognition of these heterospecific vocalizations. An intuitive explanation would be that humans use the same perceptual ability for conspecific and heterospecific discrimination, but considering that humans primarily rely on mean pitch, mean F1, and formant dispersion frequencies for conspecific voice discrimination (Baumann & Belin, 2010) but not for the discrimination of zebra finches or crows, this does not seem to be an adequate explanation. A second intuitive hypothesis would be the perception of musical patterns due to the strong influence of pitch contour. This has also not been supported by my results but would deserve more attention in future experiments. In human speech more generally, pitch contour plays an important role for the perception of intonation in speech (Hart, Cohen, & Collier, 1990) and emotional prosody (Banse & Scherer, 1996) and so the primary function of the perceptual ability used for the discrimination of zebra finch songs and crow calls may be related to either of these two areas rather than voice perception per se. Given that humans have been found to perceive the emotional state of other terrestrial vertebrates by their vocalizations (Filippi et al., 2017), emotional prosody currently appears to be a possible primary function of one of the contributing perceptual abilities. Further experiments are necessary to explore whether this is in fact a possible explanation, and whether this ability would account for the entire discrimination process or whether additional perceptual abilities are involved. However, the primary function of this perceptual ability remains speculative, and so the fitness benefits and potential selection for this ability remain speculative as well.

Future studies may also want to explore how much this perceptual abilities' tolerance for heterospecific voice cues and the structural similarities of amniote vocalizations each support individual discrimination. Is the convergent mechanism as described by the source-filter theory a factor in the achievable discrimination accuracy, or can it be achieved with vocalizations not included in this theory? Are there species that produce vocalizations in which humans cannot perceive individual differences? These questions highlight the necessity for further studies with a broader range of organisms, especially birds without pitch contour, such as flamingos or New-Zealand wrens, non-avian reptiles, species outside the amniote clade, and species outside the vertebrate clade, such as the Lusitanian toadfish (*Halobatrachus didactylus*), a vocalizing invertebrate (Vasconcelos & Ladich, 2008). The observations from the present experiments only scratch the surface of the full extent of individual acoustic discrimination and recognition of heterospecifics by humans, and much more work is needed to identify the capacities and limitations of this ability.

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

Addis, L., Friederici, A. D., Kotz, S. A., Sabisch, B., Barry, J., Richter, N., . . . Monaco, A. P. (2010). A locus for an auditory processing deficit and language impairment in an extended pedigree maps to 12p13.31-q14.3. *Genes, Brain, and Behavior*, *9*(6), 545–561. https://doi.org/10.1111/j.1601-183X.2010.00583.x

Agnolín, F. L., Egli, F. B., Chatterjee, S., Marsà, J. A. G., & Novas, F. E. (2017). Vegaviidae, a new clade of southern diving birds that survived the K/T boundary. *Die Naturwissenschaften*, *104*(11-12), 87. https://doi.org/10.1007/s00114-017-1508-y

Allen, J. B., & Neely, S. T. (1992). Micromechanical Models of the Cochlea. *Physics Today*, *45*(7), 40–47. https://doi.org/10.1063/1.881349

Anderson, J. R., Kuroshima, H., Takimoto, A., & Fujita, K. (2013). Third-party social evaluation of humans by monkeys. *Nature Communications*, *4*, 1561. https://doi.org/10.1038/ncomms2495

Arnaiz-Villena, A., Ruiz-del-Valle, V., Gomez-Prieto, P., Reguera, R., Parga-Lozano, C., & Serrano-Vela, I. (2009). Estrildinae Finches (Aves, Passeriformes) from Africa, South Asia and Australia: a Molecular Phylogeographic Study. *The Open Ornithology Journal*, *2*(1), 29–36. https://doi.org/10.2174/1874453200902010029

Audacity Team (2019). Audacity(R): Free Audio Editor and Recorder (Version 2.3.0) [Computer software]. Retrieved from https://audacityteam.org/

Aureli, F., Schaffner, C. M., Boesch, C., Bearder, S. K., Call, J., Chapman, C. A., . . . Schaik, C. P. v. (2008). Fission-Fusion Dynamics. *Current Anthropology*, *49*(4), 627–654. https://doi.org/10.1086/586708

Axelrod, R., & Hamilton, W. D. (1981). The Evolution of Cooperation. *Science*. (211), 1390–1396.

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., & Freimer, N. B. (1998). Absolute pitch: An approach for identification of genetic and nongenetic components. *American Journal of Human Genetics*, *62*(2), 224–231. https://doi.org/10.1086/301704

Bailleul, A. M., Hall, B. K., & Horner, J. R. (2012). First evidence of dinosaurian secondary cartilage in the post-hatching skull of Hypacrosaurus stebingeri (Dinosauria, Ornithischia). *PloS One*, *7*(4), e36112. https://doi.org/10.1371/journal.pone.0036112

Bamiou, D. E., Musiek, F. E., & Luxon, L. M. (2001). Aetiology and clinical presentations of auditory processing disorders--a review. *Archives of Disease in Childhood*, *85*(5), 361–365. https://doi.org/10.1136/adc.85.5.361

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Baptista, L. F., & Petrinovich, L. (1984). Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow. *Animal Behaviour*, *32*(1), 172–181. https://doi.org/10.1016/S0003-3472(84)80335-8

Baptista, L. F., & Trail, P. W. (1992). The Role of Song in the Evolution of Passerine Diversity. *Systematic Biology*, *41*(2), 242–247. https://doi.org/10.1093/sysbio/41.2.242

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

Berg, K. S., Delgado, S., Cortopassi, K. A., Beissinger, S. R., & Bradbury, J. W. (2012). Vertical transmission of learned signatures in a wild parrot. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1728), 585–591. https://doi.org/10.1098/rspb.2011.0932

Berger, L. R., & Ligon, J.D. (1977). Vocal communication and individual recognition in the pinon jay, Gymnorhinus cyanocephalus. *Animal Behaviour*, *25*, 567–584. https://doi.org/10.1016/0003-3472(77)90107-5

Boeckle, M., & Bugnyar, T. (2012). Long-term memory for affiliates in ravens. *Current Biology : CB*, *22*(9), 801–806. https://doi.org/10.1016/j.cub.2012.03.023

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer (Version 6.0.49) [Computer software]. Retrieved from http://www.praat.org/

Bossema, N. V., & Benus, R. (1985). Territorial Defense and Intra-Pair Cooperation in the Carrion Crow (Corvus-Corone). *Behavioral Ecology and Sociobiology*. (16), 99–104.

Boucaud, I. C. A., Mariette, M. M., Villain, A. S., & Vignal, C. (2016). Vocal negotiation over parental care? Acoustic communication at the nest predicts partners' incubation share. *Biological Journal of the Linnean Society*, *117*(2), 322–336. https://doi.org/10.1111/bij.12705

Bouzat, J. L. (2010). Conservation genetics of population bottlenecks: the role of chance, selection, and history. *Conservation Genetics*, *11*(2), 463–478. https://doi.org/10.1007/s10592-010-0049-0

Brewer, C. C., Zalewski, C. K., King, K. A., Zobay, O., Riley, A., Ferguson, M. A., . . . Moore, D. R. (2016). Heritability of non-speech auditory processing skills. *European Journal of Human Genetics : EJHG*, *24*(8), 1137–1144. https://doi.org/10.1038/ejhg.2015.277

Brooks, R. J., & Falls, J. B. (1975). Individual recognition by song in white-throated sparrows. I. Discrimination of songs of neighbors and strangers. *Canadian Journal of Zoology*, *53*(7), 879–888. https://doi.org/10.1139/z75-101

Brown, E. D. (1985). The Role of Song and Vocal Imitation among Common Crows (Corvus brachyrhynchos). *Zeitschrift Für Tierpsychologie*, *68*(2), 115–136. https://doi.org/10.1111/j.1439-0310.1985.tb00119.x

Brownell, W. E., & Manis, P. B. (2014). Structures, Mechanisms, and Energetics in Temporal Processing. In A. N. Popper & R. R. Fay (Eds.), *Springer Handbook of Auditory Research: Vol. 50. Perspectives on Auditory Research* (pp. 9–44). New York, NY, s.l.: Springer New York.

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2013). Voice discrimination in four primates. *Behavioural Processes*, *99*, 67–72. https://doi.org/10.1016/j.beproc.2013.06.010

Carroll, R. L. (1992). The Primary Radiation of Terrestrial Vertebrates. *Annual Review of Earth and Planetary Sciences*, *20*(1), 45–84. https://doi.org/10.1146/annurev.ea.20.050192.000401

Chamberlain, D. R., & Cornwell, G. W. (1971). Selected Vocalizations of the Common Crow. *The Auk*. (88), 613–634.

Chang, Y.-H. S., Yao, Y. [Yao], & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, *141*(2), EL120. https://doi.org/10.1121/1.4976037

Chapuis, C. (1972). Un Exemple de l'Influence du Milieu sur les Emissions Vocales des Oiseaux l'Evolution des Chants en Foret Equatoriale. *La Terre Et La Vie*. (2).

Charlton, B. D., Ellis, W. A. H., McKinnon, A. J., Cowin, G. J., Brumm, J., Nilsson, K., & Fitch, W. T. (2011). Cues to body size in the formant spacing of male koala (Phascolarctos cinereus) bellows: Honesty in an exaggerated trait. *The Journal of Experimental Biology*, *214*(Pt 20), 3414–3422. https://doi.org/10.1242/jeb.061358

Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications*, *7*, 12739. https://doi.org/10.1038/ncomms12739

Charlton, B. D., Taylor, A. M., & Reby, D. (2013). Are men better than women at acoustic size judgements? *Biology Letters*, *9*(4), 20130270. https://doi.org/10.1098/rsbl.2013.0270

Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*(3), 164–167. https://doi.org/10.1016/j.neulet.2006.06.053

Chijiiwa, H., Kuroshima, H., Hori, Y., Anderson, J. R., & Fujita, K. (2015). Dogs avoid people who behave negatively to their owner: third-party affective evaluation. *Animal Behaviour*, *106*, 123–127. https://doi.org/10.1016/j.anbehav.2015.05.018

Childers, D. G., & Lee, C. K. (1991). Vocal quality factors: Analysis, synthesis, and perception. *The Journal of the Acoustical Society of America*, *90*(5), 2394–2410. https://doi.org/10.1121/1.402044

Clark, J. A., Boersma, P., & Olmsted, D. M. (2006). Name that tune: call discrimination and individual recognition in Magellanic penguins. *Animal Behaviour*, *72*(5), 1141–1148. https://doi.org/10.1016/j.anbehav.2006.04.002

Clarke, J. A., Chatterjee, S., Li, Z. [Zhiheng], Riede, T., Agnolín, F. L., Goller, F., . . . Novas, F. E. (2016). Fossil evidence of the avian vocal organ from the Mesozoic. *Nature*, *538*(7626), 502–505. https://doi.org/10.1038/nature19852

Clarke, J. A., Tambussi, C. P., Noriega, J. I., Erickson, G. M., & Ketcham, R. A. (2005). Definitive fossil evidence for the extant avian radiation in the Cretaceous. *Nature*, *433*(7023), 305–308. https://doi.org/10.1038/nature03150

Clayton, N. S., & Emery, N. (2005). Corvid cognition. *Current Biology : CB*, *15*(3), R80-1. https://doi.org/10.1016/j.cub.2005.01.020

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, *47*(Pt 3), 207–239. https://doi.org/10.1177/00238309040470030101

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, K. M., Finney, S. C., Gibbard, P. L., & Fan, J.-X. (2013). *The ICS International Chronostratigraphic Chart*, pp. 199–204.

Colafrancesco, K. C., & Gridi-Papp, M. (2016). Vocal Sound Production and Acoustic Communication in Amphibians and Reptiles. In R. A. Suthers, W. T. Fitch, R. R. Fay, & A. N. Popper (Eds.), *Springer Handbook of Auditory Research: Vol. 53. Vertebrate Sound Production and Acoustic Communication* (pp. 51–82). Springer.

Conner, R. N. (1985). Vocalizations of Common Ravens in Virginia. *The Condor*, *87*(3), 379–388. https://doi.org/10.2307/1367219

Cornec, C., Hingrat, Y., & Rybak, F. (2014). Individual Signature in a Lekking Species: Visual and Acoustic Courtship Parameters May Help Discriminating Conspecifics in the Houbara Bustard. *Ethology*, *120*(7), 726–737. https://doi.org/10.1111/eth.12244

Cornell, H. N., Marzluff, J. M., & Pecoraro, S. (2012). Social learning spreads knowledge about dangerous humans among American crows. *Proceedings. Biological Sciences*, *279*(1728), 499–508. https://doi.org/10.1098/rspb.2011.0957

Cowan, P. J. (1974). Selective responses to the parental calls of different individual hens by young Gallus gallus: Auditory discrimination learning versus auditory imprinting. *Behavioral Biology*, *10*(4), 541–545. https://doi.org/10.1016/S0091-6773(74)92196-8

Cox, D. T. C., & Gaston, K. J. (2016). Urban Bird Feeding: Connecting People with Nature. *PloS One*, *11*(7). https://doi.org/10.1371/journal.pone.0158717

Curtin, S. (2009). Wildlife tourism: the intangible, psychological benefits of human–wildlife encounters. *Current Issues in Tourism*, *12*(5-6), 451–474. https://doi.org/10.1080/13683500903042857

Davies, N. B., & Brooke, M. D. L. (1989). An Experimental Study of Co-Evolution Between the Cuckoo, Cuculus canorus, and its Hosts. I. Host Egg Discrimination. *The Journal of Animal Ecology*, *58*(1), 207. https://doi.org/10.2307/4995

Davis, W. E. (1980). Analysis of the Vocalizations of Four Heron and Ibis Species. *Proceedings of the Colonial Waterbird Group*. (3), 216–224.

Delpey, G. (1942). Un cartilage calcifie fossile. *Bulletin De La Société Géologique De France*. (S5-XII (1-3)), 45–56.

Dooling, R. J., & Prior, N. H. (2017). Do we hear what birds hear in birdsong? *Animal Behaviour*, *124*, 283–289. https://doi.org/10.1016/j.anbehav.2016.10.012

Dufour, V., Wascher, C. A. F., Braun, A., Miller, R., & Bugnyar, T. (2012). Corvids can decide if a future exchange is worth waiting for. *Biology Letters*, *8*(2), 201–204. https://doi.org/10.1098/rsbl.2011.0726

Düring, D. N., & Elemans, C. P.H. (2016). Embodied Motor Control of Avian Vocal Production. In R. A. Suthers, W. T. Fitch, R. R. Fay, & A. N. Popper (Eds.), *Springer Handbook of Auditory Research: Vol. 53. Vertebrate Sound Production and Acoustic Communication* (pp. 119–158). Springer.

Eakle, W. L., Mannan, R. W., & Grubb, T. G. (1989). Identification of Individual Breeding Bald Eagles by Voice Analysis. *The Journal of Wildlife Management*, *53*(2), 450. https://doi.org/10.2307/3801149

El Mouden, C., West, S. A., & Gardner, A. (2010). The enforcement of cooperation by policing. *Evolution; International Journal of Organic Evolution*, *64*(7), 2139–2152. https://doi.org/10.1111/j.1558-5646.2010.00963.x

Elemans, C. P. H., Rasmussen, J. H., Herbst, C. T., Düring, D. N., Zollinger, S. A., Brumm, H., . . . Švec, J. G. (2015). Universal mechanisms of sound production and control in birds and mammals. *Nature Communications*, *6*, 8978. https://doi.org/10.1038/ncomms9978

Elie, J. E., & Theunissen, F. E. (2016). The vocal repertoire of the domesticated zebra finch: A data-driven approach to decipher the information-bearing acoustic features of communication signals. *Animal Cognition*, *19*(2), 285–315. https://doi.org/10.1007/s10071-015-0933-6

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Ericson, P. G. P., Jansén, A.-L., Johansson, U. S., & Ekman, J. (2005). Inter-generic relationships of the crows, jays, magpies and allied groups (Aves: Corvidae) based on nucleotide sequence data. *Journal of Avian Biology*, *36*(3), 222–234. https://doi.org/10.1111/j.0908-8857.2001.03409.x

Fant, G. (1960). *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.

Farquhar, C. C. (1993). Individual and Intersexual Variation in Alarm Calls of the White-Tailed Hawk. *The Condor*, *95*(1), 234–239. https://doi.org/10.2307/1369408

Farrús, M., Hernando, J., & Ejarque, P. (2007). Jitter and Shimmer Measurements for Speaker Recognition. *INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association, Antwerp, Belgium*.

Fastovsky, D. E., & Sheehan, P. M. (2005). The Extinction of the Dinosaurs in North America. *GSA Today*, *15*(3), 4. https://doi.org/10.1130/1052-5173(2005)15<4:TEOTDI>2.0.CO;2

Favaro, L., Gamba, M., Gili, C., & Pessani, D. (2017). Acoustic correlates of body size and individual identity in banded penguins. *PloS One*, *12*(2), e0170001. https://doi.org/10.1371/journal.pone.0170001

Feng, A. S., Arch, V. S., Yu, Z., Yu, X.-J., Xu, Z.-M., & Shen, J.-X. (2009). Neighbor-Stranger Discrimination in Concave-Eared Torrent Frogs, Odorrana tormota. *Ethology*, *115*(9), 851–856. https://doi.org/10.1111/j.1439-0310.2009.01666.x

Ferguson, B., & Waxman, S. R. (2016). What the beep? Six-month-olds link novel communicative signals to meaning. *Cognition*, *146*, 185–189. https://doi.org/10.1016/j.cognition.2015.09.020

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

Field, D. J., Gauthier, J. A., King, B. L., Pisani, D., Lyson, T. R., & Peterson, K. J. (2014). Toward consilience in reptile phylogeny: Mirnas support an archosaur, not lepidosaur, affinity for turtles. *Evolution & Development*, *16*(4), 189–196. https://doi.org/10.1111/ede.12081

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., . . . Güntürkün, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings. Biological Sciences*, *284*(1859). https://doi.org/10.1098/rspb.2017.0990

Fishbein, A. R., Idsardi, W. J., Ball, G. F., & Dooling, R. J. (2020). Sound sequences in birdsong: How much do birds really care? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1789), 20190044. https://doi.org/10.1098/rstb.2019.0044

Fisher, J. (1954). Evolution and bird sociality. In Huxley J., Hardy A.C., Ford E.B. (Ed.), *Evolution as a process*. London: Allen & Unwin.

Fitch, W. T. (1994). *Vocal Tract Length Perception and the Evolution of Language* (Thesis). Brown University.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2 Pt 1), 1213–1222. https://doi.org/10.1121/1.421048

Fitch, W. T. (1999). Acoustic exaggeration of size in birds via tracheal elongation: comparative and theoretical analyses. *Journal of Zoology*, *248*(1), 31–48. https://doi.org/10.1111/j.1469-7998.1999.tb01020.x

Fitch, W. T. (2009). Fossil cues to the evolution of speech. In R. P. Botha & C. Knight (Eds.), *Studies in the evolution of language: Vol. 12. The cradle of language* (pp. 112–134). Oxford, New York: Oxford University Press.

Fitch, W. T. (2016). Vertebrate Bioacoustics: Prospects and Open Problems. In R. A. Suthers, W. T. Fitch, R. R. Fay, & A. N. Popper (Eds.), *Springer Handbook of Auditory Research: Vol. 53. Vertebrate Sound Production and Acoustic Communication* (Vol. 53, pp. 297–328). Springer. https://doi.org/10.1007/978-3-319-27721-9_10

Fitch, W. T. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review*, *24*(1), 3–33. https://doi.org/10.3758/s13423-017-1236-5

Fitch, W. T. (2018). The Biology and Evolution of Speech: A Comparative Analysis. *Annual Review of Linguistics*, *4*(1), 255–279. https://doi.org/10.1146/annurev-linguistics-011817-045748

Fitch, W. T., Boer, B. de, Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, *2*(12), e1600723. https://doi.org/10.1126/sciadv.1600723

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings. Biological Sciences*, *268*(1477), 1669–1675. https://doi.org/10.1098/rspb.2001.1704

Flower, T. P., Gribble, M., & Ridley, A. R. (2014). Deception by flexible alarm mimicry in an African bird. *Science (New York, N.Y.)*, *344*(6183), 513–516. https://doi.org/10.1126/science.1249723

Ford, D. P., & Benson, R. B. J. (2020). The phylogeny of early amniotes and the affinities of Parareptilia and Varanopidae. *Nature Ecology & Evolution*, *4*(1), 57–65. https://doi.org/10.1038/s41559-019-1047-3

Forshaw, J. M., & Shephard, M. (2012). *Grassfinches in Australia*. Collingwood, VIC: CSIRO Publishing.

Forslund, P., & Larsson, K. (1991). The effect of mate change and new partner's age on reproductive success in the barnacle goose, Branta leucopsis. *Behavioral Ecology*, *2*(2), 116–122. https://doi.org/10.1093/beheco/2.2.116

Fox, J., & Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. [Computer software]. Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Francis, C. D., Ortega, C. P., & Cruz, A. (2011). Vocal frequency change reflects different responses to anthropogenic noise in two suboscine tyrant flycatchers. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1714), 2025–2031. https://doi.org/10.1098/rspb.2010.1847

Fraser, O. N., & Bugnyar, T. (2012). Reciprocity of agonistic support in ravens. *Animal Behaviour*, *83*(1), 171–177. https://doi.org/10.1016/j.anbehav.2011.10.023

Friendly, R. H., Rendall, D., & Trainor, L. J. (2013). Plasticity after perceptual narrowing for voice perception: Reinstating the ability to discriminate monkeys by their voices at 12 months of age. *Frontiers in Psychology*, *4*, 718. https://doi.org/10.3389/fpsyg.2013.00718

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

Fuong, H., Keeley, K. N., Bulut, Y., & Blumstein, D. T. (2014). Heterospecific alarm call eavesdropping in nonvocal, white-bellied copper-striped skinks, Emoia cyanura. *Animal Behaviour*, *95*, 129–135. https://doi.org/10.1016/j.anbehav.2014.07.005

Futuyma, D. J. (2013). *Evolution* (3. ed.). Sunderland, Mass.: Sinauer Associates.

Galeotti, P., & PAVAN, G. (1993). Differential responses of territorial Tawny Owls Strix aluco to the hooting of neighbours and strangers. *Ibis*, *135*(3), 300–304. https://doi.org/10.1111/j.1474-919X.1993.tb02847.x

Gardner, A. (2017). The purpose of adaptation. *Interface Focus*, *7*(5), 20170005. https://doi.org/10.1098/rsfs.2017.0005

Godefroit, P., Cau, A., Dong-Yu, H., Escuillié, F., Wenhao, W., & Dyke, G. (2013). A Jurassic avialan dinosaur from China resolves the early phylogenetic history of birds. *Nature*, *498*(7454), 359–362. https://doi.org/10.1038/nature12168

Godfrey-Smith, P. (2007). Conditions for Evolution by Natural Selection. *Journal of Philosophy*, *104*(10), 489–516. https://doi.org/10.5840/jphil2007104103

Goodale, E., & Kotagama, S. W. (2008). Response to conspecific and heterospecific alarm calls in mixed-species bird flocks of a Sri Lankan rainforest. *Behavioral Ecology*, *19*(4), 887–894. https://doi.org/10.1093/beheco/arn045

Goodwin, D. (1976). *Crows of the world*. Ithaca, N.Y.: Cornell Univ. Press.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383–406. https://doi.org/10.1006/jpho.2001.0147

Grava, T., Mathevon, N., Place, E., & Balluet, P. (2008). Individual acoustic monitoring of the European Eagle Owl Bubo bubo. *Ibis*, *150*(2), 279–287. https://doi.org/10.1111/j.1474-919X.2007.00776.x

Guillette, L. M., Hoeschele, M., Hahn, A. H., & Sturdy, C. B. (2013). Heterospecific discrimination of Poecile vocalizations by zebra finches (Taeniopygia guttata). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *127*(3), 227–236. https://doi.org/10.1037/a0029992

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biology*, *5*(12), e321. https://doi.org/10.1371/journal.pbio.0050321

Hamilton, W. D. (1964). The Genetical Evolution of Social Behaviour. I. *Journal of Theoretical Biology*. (7), 1–16.

Hamilton, W. D. (1972). Altruism and Related Phenomena, Mainly in Social Insects. *Annual Review of Ecology and Systematics*, *3*(1), 193–232. https://doi.org/10.1146/annurev.es.03.110172.001205

Haring, E., Däubl, B., Pinsker, W., Kryukov, A., & Gamauf, A. (2012). Genetic divergences and intraspecific variation in corvids of the genus Corvus (Aves: Passeriformes: Corvidae) - a first survey based on museum specimens. *Journal of Zoological Systematics and Evolutionary Research*, *50*(3), 230–246. https://doi.org/10.1111/j.1439-0469.2012.00664.x

Harrington, J., & Cassidy, S. (1999). *Techniques in Speech Acoustics*. *Text, Speech and Language Technology: Vol. 8*. Dordrecht: Springer. https://doi.org/10.1007/978-94-011-4657-9

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

Harrison, C. J. O., Nicolai, J., Immelmann, K., & Wolters, H. E. (1962). Solitary Song and its inhibition in some Estrildidae. *Journal of Ornithology*, *103*(4), 369–379. https://doi.org/10.1007/BF01676599

Hart, J. 't, Cohen, A., & Collier, R. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. *Cambridge studies in speech science and communication*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627743

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science (New York, N.Y.)*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy*, *40*(3), 217–234. https://doi.org/10.1016/S0005-7967(01)00006-7

Hirano, M., Ohala, J., & Vennard, W. (1969). The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. *Journal of Speech and Hearing Research*, *12*(3), 616–628. https://doi.org/10.1044/jshr.1203.616

Hodges, S. A. (2012). *Routes to cooperation in the rifleman Acanthisitta chloris* (Thesis (Ph.D.)). University of Sheffield.

Hogg, D. A. (1982). Ossification of the laryngeal, tracheal and syringeal cartilages in the domestic fowl. *Journal of Anatomy*, *134*(Pt 1), 57–71.

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research*, *366*, 50–64. https://doi.org/10.1016/j.heares.2018.06.014

Hong, Y. T., Hong, K. H., Jun, J.-P., & Hwang, P. H. (2015). The effects of dynamic laryngeal movements on pitch control. *American Journal of Otolaryngology*, *36*(5), 660–665. https://doi.org/10.1016/j.amjoto.2015.05.002

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. R. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *The Journal* *of the Acoustical Society of America*, *106*(3 Pt 1), 1532–1542. https://doi.org/10.1121/1.427150

Hudspeth, A. J., & Lewis, R. S. (1988). A model for electrical resonance and frequency tuning in saccular hair cells of the bull-frog, Rana catesbeiana. *The Journal of Physiology*, *400*, 275–297. https://doi.org/10.1113/jphysiol.1988.sp017120

Illmann, G., Špinka, M., Schrader, L., & Šustr, P. (2002). Acoustical mother-offspring recognition in pigs (sus scrofa domestica). *Behaviour*, *139*(4), 487–505. https://doi.org/10.1163/15685390260135970

Ito, R., & Mori, A. (2010). Vigilance against predators induced by eavesdropping on heterospecific alarm calls in a non-vocal lizard Oplurus cuvieri cuvieri (Reptilia: Iguania). *Proceedings. Biological Sciences*, *277*(1685), 1275–1280. https://doi.org/10.1098/rspb.2009.2047

Izawa, E.-I., & Watanabe, S. (2008). Formation of linear dominance relationship in captive jungle crows (Corvus macrorhynchos): Implications for individual recognition. *Behavioural Processes*, *78*(1), 44–52. https://doi.org/10.1016/j.beproc.2007.12.010

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, *366*(6461), 50–54. https://doi.org/10.1126/science.aax0287

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., . . . Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science (New York, N.Y.)*, *346*(6215), 1320–1331. https://doi.org/10.1126/science.1253451

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, *24*(10), 1995–2004. https://doi.org/10.1177/0956797613482467

Jønsson, K. A., Fabre, P.-H., & Irestedt, M. (2012). Brains, tools, innovation and biogeography in crows and ravens. *BMC Evolutionary Biology*, *12*, 72. https://doi.org/10.1186/1471-2148-12-72

Jouventin, Aubin, T., & Lengagne (1999). Finding a parent in a king penguin colony: The acoustic system of individual recognition. *Animal Behaviour*, *57*(6), 1175–1183. https://doi.org/10.1006/anbe.1999.1086

Khwaja, N., Briskie, J. V., & Hatchwell, B. J. (2019). Individuality, kin similarity and experimental playback of contact calls in cooperatively breeding riflemen. *New Zealand Journal of Zoology*, *46*(4), 334–347. https://doi.org/10.1080/03014223.2019.1587477

Kim, J. (2019). Individual differences in the production of prosodic boundaries in American English. *Physics Today*, *145*(3), 1933. https://doi.org/10.1121/1.5102039

Kingsley, E. P., Eliason, C. M., Riede, T., Li, Z. [Zhiheng], Hiscock, T. W., Farnsworth, M., . . . Clarke, J. A. (2018). Identity and novelty in the avian syrinx. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(41), 10209–10217. https://doi.org/10.1073/pnas.1804586115

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., . . . Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224. https://doi.org/10.1111/1467-9280.02435

Klenova, A. V., Volodin, I. A., & Volodina, E. V. (2009). The variation in reliability of individual vocal signature throughout ontogenesis in the red-crowned crane Grus japonensis. *Acta Ethologica*, *12*(1), 29–36. https://doi.org/10.1007/s10211-009-0053-x

Knoblauch, K. (2014). psyphy: Functions for analyzing psychophysical data in R (Version 0.1-9) [Computer software]. Retrieved from https://CRAN.R-project.org/package=psyphy

Kondo, N., & Hiraiwa-Hasegawa, M. (2015). The influence of social dominance on calling rate in the Large-billed Crow (Corvus macrorhynchos). *Journal of Ornithology*, *156*(3), 775–782. https://doi.org/10.1007/s10336-015-1191-8

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

Kondo, N., Izawa, E.-I., & Watanabe, S. (2012). Crows cross-modally recognize group members but not non-group members. *Proceedings. Biological Sciences*, *279*(1735), 1937–1942. https://doi.org/10.1098/rspb.2011.2419

Kondo, N., Watanabe, S., & Izawa, E.-I. (2010). A Temporal Rule in Vocal Exchange Among Large-Billed Crows Corvus macrorhynchos in Japan. *Ornithological Science*, *9*(1), 83–91. https://doi.org/10.2326/osj.9.83

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

Krull, C. R., Parsons, S., & Hauber, M. E. (2009). The presence of ultrasonic harmonics in the calls of the rifleman (Acanthisitta chloris). *Notornis*. (56), 158–161.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science (New York, N.Y.)*, *190*(4209), 69–72. https://doi.org/10.1126/science.1166301

Kümmerli, R., Gardner, A., West, S. A., & Griffin, A. S. (2009). Limited dispersal, budding dispersal, and cooperation: An experimental study. *Evolution; International Journal of Organic Evolution*, *63*(4), 939–949. https://doi.org/10.1111/j.1558-5646.2008.00548.x

La Padilla de Torre, M., Briefer, E. F., Ochocki, B. M., McElligott, A. G., & Reader, T. (2016). Mother–offspring recognition via contact calls in cattle, Bos taurus. *Animal Behaviour*, *114*, 147–154. https://doi.org/10.1016/j.anbehav.2016.02.004

Lachlan, R. F., van Heijningen, C. A. A., Ter Haar, S. M., & ten Cate, C. (2016). Zebra Finch Song Phonology and Syntactical Structure across Populations and Continents-A Computational Comparison. *Frontiers in Psychology*, *7*, 980. https://doi.org/10.3389/fpsyg.2016.00980

Ladich, F., & Winkler, H. (2017). Acoustic communication in terrestrial and aquatic vertebrates. *The Journal of Experimental Biology*, *220*(Pt 13), 2306–2317. https://doi.org/10.1242/jeb.132944

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*(6855), 519–523. https://doi.org/10.1038/35097076

Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A., & Shumaker, R. W. (2016). Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific Reports*, *6*, 30315. https://doi.org/10.1038/srep30315

Lange-Küttner, C. (2010). Discrimination of sea-bird sounds vs. garden-bird songs: Do Scottish and German-Saxon infants show the same preferential looking behaviour as adults? *European Journal of Developmental Psychology*, *7*(5), 578–602. https://doi.org/10.1080/17405620902937531

Lea, A. J., Barrera, J. P., Tom, L. M., & Blumstein, D. T. (2008). Heterospecific eavesdropping in a nonsocial species. *Behavioral Ecology*, *19*(5), 1041–1046. https://doi.org/10.1093/beheco/arn064

Leroux, M., Hetem, R. S., Hausberger, M., & Lemasson, A. (2018). Cheetahs discriminate familiar and unfamiliar human voices. *Scientific Reports*, *8*(1), 15516. https://doi.org/10.1038/s41598-018-33971-1

Lesbarrères, D., & Lodé, T. (2002). Variations in male calls and responses to an unfamiliar advertisement call in a territorial breeding anuran, Rana dalmatina : evidence for a “dear enemy” effect. *Ethology Ecology & Evolution*, *14*(4), 287–295. https://doi.org/10.1080/08927014.2002.9522731

Lessells, C. M., Rowe, C. L., & McGregor, P. K. (1995). Individual and sex differences in the provisioning calls of European bee-eaters. *Animal Behaviour*, *49*(1), 244–247. https://doi.org/10.1016/0003-3472(95)80174-X

Li, Z. [Zhiheng], Zhou, Z., & Clarke, J. A. (2018). Convergent evolution of a mobile bony tongue in flighted dinosaurs and pterosaurs. *PloS One*, *13*(6), e0198078. https://doi.org/10.1371/journal.pone.0198078

Lloyd-Jones, D. J. (2014). Ultrasonic harmonics in the calls of rock wren (Xenicus gilviventris). *Notornis*. (61), 165–169.

Longrich, N. R., Bhullar, B.-A. S., & Gauthier, J. A. (2012). Mass extinction of lizards and snakes at the Cretaceous-Paleogene boundary. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(52), 21396–21401. https://doi.org/10.1073/pnas.1211526110

Longrich, N. R., Scriberas, J., & Wills, M. A. (2016). Severe extinction and rapid recovery of mammals across the Cretaceous-Palaeogene boundary, and the effects of rarity on patterns of extinction and recovery. *Journal of Evolutionary Biology*, *29*(8), 1495–1512. https://doi.org/10.1111/jeb.12882

Longrich, N. R., Tokaryk, T., & Field, D. J. (2011). Mass extinction of birds at the Cretaceous-Paleogene (K-Pg) boundary. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(37), 15253–15257. https://doi.org/10.1073/pnas.1110395108

Lopez-Poveda, E. A. (2014). Development of Fundamental Aspects of Human Auditory Perception. In R. Romand & I. Varela-Nieto (Eds.), *Development of Auditory and Vestibular Systems*. Academic Press.

Lorenz, K. (1952). *King Solomon's ring. New light on animal ways*. London: Routledge (Routledge classics).

Mager III, J. N., Walcott, C., & Piper, W. H. (2010). Common Loons can differentiate yodels of neighboring and non-neighboring conspecifics. *Journal of Field Ornithology*, *81*(4), 392–401. https://doi.org/10.1111/j.1557-9263.2010.00295.x

Magrath, R. D., Haff, T. M., McLachlan, J. R., & Igic, B. (2015). Wild birds learn to eavesdrop on heterospecific alarm calls. *Current Biology : CB*, *25*(15), 2047–2050. https://doi.org/10.1016/j.cub.2015.06.028

Manley, G. A. (2000). Cochlear mechanisms from a phylogenetic viewpoint. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11736–11743. https://doi.org/10.1073/pnas.97.22.11736

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings. Biological Sciences*, *268*(1483), 2315–2324. https://doi.org/10.1098/rspb.2001.1773

Marsà, J. A. G., Agnolín, F. L., & Novas, F. (2019). Bone microstructure of Vegavis iaai (Aves, Anseriformes) from the Upper Cretaceous of Vega Island, Antarctic Peninsula. *Historical Biology*, *31*(2), 163–167. https://doi.org/10.1080/08912963.2017.1348503

Marzluff, J. M., & Angell, T. (2012). *Gifts of the Crow: How Perception, Emotion, and Thought Allow Smart Birds to Behave Like Huamsn*. New York: Atria Paperback.

Masin, S., Massa, R., & Bottoni, L. (2004). Evidence of tutoring in the development of subsong in newly-fledged Meyer's Parrots Poicephalus meyeri. *Anais Da Academia Brasileira De Ciencias*, *76*(2), 231–236. https://doi.org/10.1590/s0001-37652004000200006

Massen, J. J.M., Ritter, C., & Bugnyar, T. (2015). Tolerance and reward equity predict cooperation in ravens (Corvus corax). *Scientific Reports*, *5*, 15021. https://doi.org/10.1038/srep15021

Mates, E. A., Tarter, R. R., Ha, J. C., Clark, A. B., & McGowan, K. J. (2015). Acoustic profiling in a complexly social species, the American crow: Caws encode information on caller sex, identity, and behavioural context. *Bioacoustics*, *24*(1), 63–80. https://doi.org/10.1080/09524622.2014.933446

Mathevon, N. (1996). What parameters can be used for individual acoustic recognition by the greater flamingo? *Comptes Rendus De L'academie Des Sciences. Serie III, Sciences De La Vie*, *319*(1), 29–32.

Matsubara, H. (2007). Large-billed Crow Hashibuto-garasu (Jpn) Corvus macrorhynchos. *Bird Research News*. (4).

Matsumoto, H., Hiki, S., Sone, T., & Nimura, T. (1973). Multidimensional representation of personal quality of vowels and its acoustical correlates. *IEEE Transactions on Audio and Electroacoustics*, *21*(5), 428–436. https://doi.org/10.1109/TAU.1973.1162507

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge Univ. Press.

McComb, Moss, Sayialel, & Baker (2000). Unusually extensive networks of vocal recognition in African elephants. *Animal Behaviour*, *59*(6), 1103–1109. https://doi.org/10.1006/anbe.2000.1406

McLachlan, S. M. S., Kaiser, G. W., & Longrich, N. R. (2017). Maaqwi cascadensis: A large, marine diving bird (Avialae: Ornithurae) from the Upper Cretaceous of British Columbia, Canada. *PloS One*, *12*(12), e0189473. https://doi.org/10.1371/journal.pone.0189473

McLeod, A. I. (2011). Kendall: Kendall rank correlation and Mann-Kendall trend test (Version 2.2) [Computer software]. Retrieved from https://CRAN.R-project.org/package=Kendall

Mérő, L. (1998). The Prisoner's Dilemma. In L. Mero (Ed.), *Moral Calculations: Game Theory, Logic, and Human Frailty* (pp. 28–47). New York, NY: Springer. https://doi.org/10.1007/978-1-4612-1654-4_3

Miller, D. B. (1979). Long-term recognition of father's song by female zebra finches. *Nature*, *280*(5721), 389–391. https://doi.org/10.1038/280389a0

Müller, J.J.A., Massen, J. J.M., Bugnyar, T., & Osvath, M. (2017). Ravens remember the nature of a single reciprocal interaction sequence over 2 days and even after a month. *Animal Behaviour*, *128*, 69–78. https://doi.org/10.1016/j.anbehav.2017.04.004

Naoi, N., Watanabe, S., Maekawa, K., & Hibiya, J. (2012). Prosody discrimination by songbirds (Padda oryzivora). *PloS One*, *7*(10), e47446. https://doi.org/10.1371/journal.pone.0047446

Norton, P., Barschke, P., Scharff, C., & Mendoza, E. (2019). Differential song deficits after lentivirus-mediated knockdown of FoxP1, FoxP2 or FoxP4 in Area X of juvenile zebra finches. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience.* Advance online publication. https://doi.org/10.1523/JNEUROSCI.1250-19.2019

Nottebohm, F. (1975). Continental Patterns of Song Variability in Zonotrichia capensis: Some Possible Ecological Correlates. *The American Naturalist*, *109*(970), 605–624. https://doi.org/10.1086/283033

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291–1298. https://doi.org/10.1038/nature04131

Nuttall, R. J. (1993). VOCAL BEHAVIOUR OF THE QUAIL FINCH ORTYGOSPIZA ATRICOLLIS. *Ostrich*, *64*(3), 97–104. https://doi.org/10.1080/00306525.1993.9632641

Obozova, T. A., Smirnova, A. A., & Zorina, Z. A. (2018). Early Ontogeny of the Behavior of Young in Large-Billed Crows (Corvus macrorhynchos) in Their Natural Habitat. *Biology Bulletin*, *45*(8), 794–802. https://doi.org/10.1134/S1062359018080125

Ockleford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of neonates to parents' and others' voices. *Early Human Development*, *18*(1), 27–36. https://doi.org/10.1016/0378-3782(88)90040-0

Oghalai, J. S. (2004). The cochlear amplifier: Augmentation of the traveling wave within the inner ear. *Current Opinion in Otolaryngology & Head and Neck Surgery*, *12*(5), 431–438.

Ohms, V. R., Escudero, P., Lammers, K., & ten Cate, C. (2012). Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception. *Animal Cognition*, *15*(2), 155–161. https://doi.org/10.1007/s10071-011-0441-2

Ohms, V. R., Gill, A., van Heijningen, C. A. A., Beckers, G. J. L., & ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings. Biological Sciences*, *277*(1684), 1003–1009. https://doi.org/10.1098/rspb.2009.1788

Ohms, V. R., Snelderwaard, P. C., ten Cate, C., & Beckers, G. J. L. (2010). Vocal tract articulation in zebra finches. *PloS One*, *5*(7), e11923. https://doi.org/10.1371/journal.pone.0011923

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733-45. https://doi.org/10.1093/nar/gkv1189

Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., . . . Faircloth, B. C. (2019). Earth history and the passerine superradiation. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(16), 7916–7925. https://doi.org/10.1073/pnas.1813206116

Ostojić, L., & Clayton, N. S. (2014). Behavioural coordination of dogs in a cooperative problem-solving task with a conspecific and a human partner. *Animal Cognition*, *17*(2), 445–459. https://doi.org/10.1007/s10071-013-0676-1

Ou, J., & Law, S.-P. (2017). Cognitive basis of individual differences in speech perception, production and representations: The role of domain general attentional switching. *Attention, Perception & Psychophysics*, *79*(3), 945–963. https://doi.org/10.3758/s13414-017-1283-z

Päckert, M. (2018). Song: The Learned Language of Three Major Bird Clades. In D. T. Tietze (Ed.), *Fascinating Life Sciences. Bird Species: How They Arise, Modify and Vanish* (Vol. 511, pp. 75–94). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-91689-7_5

Padget, A. E. (2010). *Structure and possible function of vocalizations of captive red-legged seriemas (Cariama cristata)* (Master's thesis). University of Tennessee.

Paradis E., & Schliep K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. (35), 526–528.

Paul, G. S. (2010). *The Princeton field guide to dinosaurs*. *Princeton field guides*. Princeton: Princeton University Press.

Payne, R. B. (1981). Song learning and social interaction in indigo buntings. *Animal Behaviour*, *29*(3), 688–697. https://doi.org/10.1016/S0003-3472(81)80003-6

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., . . . 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

Perry, J. C., & Rowe, L. (2015). The evolution of sexually antagonistic phenotypes. *Cold Spring Harbor Perspectives in Biology*, *7*(6). https://doi.org/10.1101/cshperspect.a017558

Perszyk, D. R., & Waxman, S. R. (2016). Listening to the calls of the wild: The role of experience in linking language and cognition in young infants. *Cognition*, *153*, 175–181. https://doi.org/10.1016/j.cognition.2016.05.004

Petrie, M., & Kempenaers, B. (1998). Extra-pair paternity in birds: explaining variation between species and populations. *Trends in Ecology & Evolution*, *13*(2), 52–58. https://doi.org/10.1016/S0169-5347(97)01232-9

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., . . . Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science (New York, N.Y.)*, *346*(6215), 1256846. https://doi.org/10.1126/science.1256846

Pike, C. D., & Kriengwatana, B. P. (2019). Vocal tract constancy in birds and humans. *Behavioural Processes*, *163*, 99–112. https://doi.org/10.1016/j.beproc.2018.08.001

Pisanski, K., Raine, J., & Reby, D. (2020). Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries. *Royal Society Open Science*, *7*(2), 191642. https://doi.org/10.1098/rsos.191642

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

Pitcher, B. J., Briefer, E. F., Vannoni, E., & McElligott, A. G. (2014). Fallow bucks attend to vocal cues of motivation and fatigue. *Behavioral Ecology*, *25*(2), 392–401. https://doi.org/10.1093/beheco/art131

Pitcher, B. J., Harcourt, R. G., & Charrier, I. (2010). The memory remains: Long-term vocal recognition in Australian sea lions. *Animal Cognition*, *13*(5), 771–776. https://doi.org/10.1007/s10071-010-0322-0

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 976–986. https://doi.org/10.1037//0096-1523.20.5.976

Policht, R., Petru, M., Lastimoza, L., & Suarez, L. E.O. (2009). Potential for the use of vocal individuality as a conservation research tool in two threatened Philippine hornbill species, the Visayan Hornbill and the Rufous-headed Hornbill. *Bird Conservation International*, *19*(1), 83–97. https://doi.org/10.1017/S0959270908008228

Popham, S., Boebinger, D., Ellis, D. P. W., Kawahara, H., & McDermott, J. H. (2018). Inharmonic speech reveals the role of harmonicity in the cocktail party problem. *Nature Communications*, *9*(1), 2122. https://doi.org/10.1038/s41467-018-04551-8

Prior, N. H., Smith, E., Lawson, S., Ball, G. F., & Dooling, R. J. (2018). Acoustic fine structure may encode biologically relevant information for zebra finches. *Scientific Reports*, *8*(1), 6212. https://doi.org/10.1038/s41598-018-24307-0

Proops, L., McComb, K., & Reby, D. (2009). Cross-modal individual recognition in domestic horses (Equus caballus). *Proceedings of the National Academy of Sciences of the United States of America*, *106*(3), 947–951. https://doi.org/10.1073/pnas.0809127105

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*(7574), 569–573. https://doi.org/10.1038/nature15697

Qi, Y., & Hillman, R. E. (1997). Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *Physics Today*, *102*(1), 537–543. https://doi.org/10.1121/1.419726

R Core Team (2019). R: A Language and Environment for Statistical Computing, Vienna, Austria (Version 3.6.1) [Computer software]. Retrieved from https://www.R-project.org/

Ramakrishnan, U., & Coss, R. G. (2000). Recognition of heterospecific alarm vocalizations by bonnet macaques (Macaca radiata). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *114*(1), 3–12.

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

Range, F., Kassis, A., Taborsky, M., Boada, M., & Marshall-Pescini, S. (2019). Wolves and dogs recruit human partners in the cooperative string-pulling task. *Scientific Reports*, *9*(1), 17591. https://doi.org/10.1038/s41598-019-53632-1

Raup, D. M. (1972). Taxonomic Diversity during the Phanerozoic. *Science (New York, N.Y.)*, *177*(4054), 1065–1071. https://doi.org/10.1126/science.177.4054.1065

Raymond, S., Spotswood, S., Clarke, H., Zielonka, N., Lowe, A., & Durrant, K. L. (2019). Vocal instability over time in individual male European nightjars, Caprimulgus europaeus : recommendations for acoustic monitoring and surveys. *Bioacoustics*, *72*(1), 1–16. https://doi.org/10.1080/09524622.2019.1603121

Reber, S. A., Nishimura, T., Janisch, J., Robertson, M., & Fitch, W. T. (2015). A Chinese alligator in heliox: Formant frequencies in a crocodilian. *The Journal of Experimental Biology*, *218*(Pt 15), 2442–2447. https://doi.org/10.1242/jeb.119552

Reby, D., McComb, K., Darwin, C., & Fitch, W. T. (2005). Formant frequencies as indicators of body size in red deer roars. *Physics Today*, *117*(4), 2373. https://doi.org/10.1121/1.4785616

Remage-Healey, L., Nowacek, D. P., & Bass, A. H. (2006). Dolphin foraging sounds suppress calling and elevate stress hormone levels in a prey species, the Gulf toadfish. *The Journal of Experimental Biology*, *209*(Pt 22), 4444–4451. https://doi.org/10.1242/jeb.02525

Rendall, D., Owren, M. J., & Rodman, P. S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (Macaca mulatta) vocalizations. *Physics Today*, *103*(1), 602–614. https://doi.org/10.1121/1.421104

Rice, M. L., & Smolík, F. (2011). Genetics of language disorders: clinical conditions, phenotypes, and genes. In M. G. Gaskell & G. T. M. Altmann (Eds.), *The Oxford handbook of psycholinguistics* (pp. 684–700). Oxford [England], New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0042

Riede, T., & Brown, C. (2013). Body Size, Vocal Fold Length, and Fundamental Frequency – Implications for Mammal Vocal Communication. *Nova Acta Leopoldina: Abhandlungen Der Kaiserlich Leopoldinisch-Carolinisch Deutschen Akademie Der Naturforscher*. (111), 295–314.

Riede, T., Eliason, C. M., Miller, E. H., Goller, F., & Clarke, J. A. (2016). Coos, booms, and hoots: The evolution of closed-mouth vocal behavior in birds. *Evolution; International Journal of Organic Evolution*, *70*(8), 1734–1746. https://doi.org/10.1111/evo.12988

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

Riede, T., & Goller, F. (2014). Morphological basis for the evolution of acoustic diversity in oscine songbirds. *Proceedings. Biological Sciences*, *281*(1779), 20132306. https://doi.org/10.1098/rspb.2013.2306

Riede, T., Thomson, S. L., Titze, I. R., & Goller, F. (2019). The evolution of the syrinx: An acoustic theory. *PLoS Biology*, *17*(2), e2006507. https://doi.org/10.1371/journal.pbio.2006507

Root-Gutteridge, H., Bencsik, M., Chebli, M., Gentle, L. K., Terrell-Nield, C., Bourit, A., & Yarnell, R. W. (2014). Identifying individual wild Eastern grey wolves (Canis lupus lycaon) using fundamental frequency and amplitude of howls. *Bioacoustics*, *23*(1), 55–66. https://doi.org/10.1080/09524622.2013.817317

Sacchi, R., Galeotti, P., Fasola, M., & Gerzeli, G. (2004). Larynx morphology and sound production in three species of Testudinidae. *Journal of Morphology*, *261*(2), 175–183. https://doi.org/10.1002/jmor.10236

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., & Izawa, E.-I. (2020). Language Discrimination by Large-Billed Crows. In Ravignani, A., Barbieri, C., Martins, M., Flaherty, M., Jadoul, Y., Lattenkamp, E., Little, H., Mudd, K. & Verhoef, T. (Chair), *EvoLang13,* Brussels, Belgium.

Schulte, P., Alegret, L., Arenillas, I., Arz, J. A., Barton, P. J., Bown, P. R., . . . Willumsen, P. S. (2010). The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science (New York, N.Y.)*, *327*(5970), 1214–1218. https://doi.org/10.1126/science.1177265

Schweinfurth, M. K., & Taborsky, M. (2020). Rats play tit-for-tat instead of integrating social experience over multiple interactions. *Proceedings. Biological Sciences*, *287*(1918), 20192423. https://doi.org/10.1098/rspb.2019.2423

Searby, A., & Jouventin, P. (2003). Mother-lamb acoustic recognition in sheep: A frequency coding. *Proceedings. Biological Sciences*, *270*(1526), 1765–1771. https://doi.org/10.1098/rspb.2003.2442

Sharp, S. P., McGowan, A., Wood, M. J., & Hatchwell, B. J. (2005). Learned kin recognition cues in a social bird. *Nature*, *434*(7037), 1127–1130. https://doi.org/10.1038/nature03522

Shriner (1998). Yellow-bellied marmot and golden-mantled ground squirrel responses to heterospecific alarm calls. *Animal Behaviour*, *55*(3), 529–536.

Sierro, J., Loretto, M.-C., Szipl, G., Massen, J. J.M., & Bugnyar, T. (2019). Food calling in wild ravens (Corvus corax) revisited: Who is addressed? *Ethology*, *70*(2), 305. https://doi.org/10.1111/eth.12991

Slabbekoorn, H., & den Boer-Visser, A. (2006). Cities change the songs of birds. *Current Biology*, *16*(23), 2326–2331. https://doi.org/10.1016/j.cub.2006.10.008

Slater, P.J.B., & Clayton, N. S. (1991). Domestication and Song Learning in Zebra Finches Taeniopygia guttata. *Emu - Austral Ornithology*, *91*(2), 126–128. https://doi.org/10.1071/MU9910126

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

Soberon Mainero, J., & Martinez del Rio, C. (1985). Cheating and Taking Advantage in Mutualistic Assosciations. In D. H. Boucher (Ed.), *The biology of mutualism: Ecology and evolution* (pp. 192–213). New York: Oxford Univ. Press.

Soma, M., & Hasegawa, T. (2003). Formation of foraging flocks with recruitment calls in Jungle Crows, Corvus macrorhynchos. *Japanese Journal of Ornithology*, *52*(2), 97–106. https://doi.org/10.3838/jjo.52.97

Sorcinelli, A., Ference, J., Curtin, S., & Vouloumanos, A. (2019). Preference for speech in infancy differentially predicts language skills and autism-like behaviors. *Journal of Experimental Child Psychology*, *178*, 295–316. https://doi.org/10.1016/j.jecp.2018.09.011

Sossinka, R., & Böhner, J. (1980). Song Types in the Zebra Finch Poephila guttata castanotis1. *Zeitschrift Für Tierpsychologie*, *53*(2), 123–132. https://doi.org/10.1111/j.1439-0310.1980.tb01044.x

Spierings, M. J., & ten Cate, C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1787). https://doi.org/10.1098/rspb.2014.0480

Spottiswoode, C. N., Begg, K. S., & Begg, C. M. (2017). *Data from: Reciprocal signaling in honeyguide-human mutualism*. https://doi.org/10.5061/DRYAD.8C65S

Sridhar, H., & Guttal, V. (2018). Friendship across species borders: Factors that facilitate and constrain heterospecific sociality. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1746). https://doi.org/10.1098/rstb.2017.0014

Stahler, D., Heinrich, B., & Smith, D. [Douglas] (2002). Common ravens, Corvus corax, preferentially associate with grey wolves, Canis lupus, as a foraging strategy in winter. *Animal Behaviour*, *64*(2), 283–290. https://doi.org/10.1006/anbe.2002.3047

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

Stephens, C. (1996). Modelling Reciprocal Altruism. *The British Journal for the Philosophy of Science*, *47*(4), 533–551. https://doi.org/10.1093/bjps/47.4.533

Stoeger, A. S., Mietchen, D., Oh, S., Silva, S. de, Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology : CB*, *22*(22), 2144–2148. https://doi.org/10.1016/j.cub.2012.09.022

St-Pierre, A., Larose, K., & Dubois, F. (2009). Long-term social bonds promote cooperation in the iterated Prisoner's Dilemma. *Proceedings. Biological Sciences*, *276*(1676), 4223–4228. https://doi.org/10.1098/rspb.2009.1156

Sung, H.-C., & Miller, E. H. (2007). Variation and Individuality in Display-Flight Calls of Male Piping Plovers. *The Condor*, *109*(2), 468–475. https://doi.org/10.1093/condor/109.2.468

Swaddle, J. P. (2010). The zebra finch. In M. Breed & J. Moore (Eds.), *The encyclopedia of animal behavior*. Elsevier.

Szipl, G., Boeckle, M., Werner, S. A. B., & Kotrschal, K. (2014). Mate recognition and expression of affective state in croop calls of Northern Bald Ibis (Geronticus eremita). *PloS One*, *9*(2), e88265. https://doi.org/10.1371/journal.pone.0088265

Takagi, M. (2020). Vocalizations of the Ryukyu Scops Owl Otus elegans : individually recognizable and stable. *Bioacoustics*, *29*(1), 28–44. https://doi.org/10.1080/09524622.2018.1539925

Takahasi, M., Yamada, H., & Okanoya, K. (2010). Statistical and Prosodic Cues for Song Segmentation Learning by Bengalese Finches (Lonchura striata var. domestica). *Ethology*, *116*(6), 481–489. https://doi.org/10.1111/j.1439-0310.2010.01772.x

Takeuchi, A. H., & Hulse, S. H. (1991). Absolute-Pitch Judgments of Black and White-Key Pitches. *Music Perception: An Interdisciplinary Journal*, *9*(1), 27–46. https://doi.org/10.2307/40286157

Tanalgo, K., Waldien, D., Monfort, N., & Hughes, A. (2019). *Attacked from above and below, new observations of cooperative and solitary predators on roosting cave bats* (Vol. 115). https://doi.org/10.1101/550582

Tarter, R. R. (2008). *The Vocal Behavior of the American Crow, Corvus brachyrhyncos.* Retrieved from Retrieved from https://etd.ohiolink.edu/

Taylor, A. M., Charlton, B. D., & Reby, D. (2016). Vocal Production by Terrestrial Mammals: Source, Filter, and Function. In R. A. Suthers, W. T. Fitch, R. R. Fay, & A. N. Popper (Eds.), *Springer Handbook of Auditory Research: Vol. 53. Vertebrate Sound Production and Acoustic Communication* (pp. 229–260). Springer.

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

Taylor, A. M., Reby, D., & McComb, K. (2009). Context-Related Variation in the Vocal Growling Behaviour of the Domestic Dog (Canis familiaris). *Ethology*, *115*(10), 905–915. https://doi.org/10.1111/j.1439-0310.2009.01681.x

Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H., & White, S. A. (2004). Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *24*(13), 3152–3163. https://doi.org/10.1523/JNEUROSCI.5589-03.2004

Thomas, B. A., & Cleal, C. J. (2018). Arborescent lycophyte growth in the late Carboniferous coal swamps. *The New Phytologist*, *218*(3), 885–890. https://doi.org/10.1111/nph.14903

Thompson, N. S. (1982). A Comparison of Cawing in the European Carrion Crow (Corvus Corone) and the American Common Crow (Corvus Brachyrhynchos). *Behaviour*, *80*(1-2), 106–117. https://doi.org/10.1163/156853982X00472

Tibbetts, E. A., & Dale, J. (2007). Individual recognition: It is good to be different. *Trends in Ecology & Evolution*, *22*(10), 529–537. https://doi.org/10.1016/j.tree.2007.09.001

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

Tomback, D. F. (1986). Observations on the Behavior and Ecology of the Mariana Crow. *The Condor*, *88*(3), 398–401. https://doi.org/10.2307/1368898

Tomić, N., & Meyer-Rochow, V. B. (2011). Atavisms: Medical, genetic, and evolutionary implications. *Perspectives in Biology and Medicine*, *54*(3), 332–353. https://doi.org/10.1353/pbm.2011.0034

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

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

Townsend, S., Charlton, B. D., & Manser, M. B. (2014). Acoustic cues to identity and predator context in meerkat barks. *Animal Behaviour*. (94), 143–149.

Trefry, S. A., & Hik, D. S. (2009). Eavesdropping on the Neighbourhood: Collared Pika (Ochotona collaris) Responses to Playback Calls of Conspecifics and Heterospecifics. *Ethology*, *115*(10), 928–938. https://doi.org/10.1111/j.1439-0310.2009.01675.x

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, *46*(1), 35–57. https://doi.org/10.1086/406755

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38 (Pt 3)*, 253–265. https://doi.org/10.1177/002383099503800303

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, *25*(5), 829–834. https://doi.org/10.1016/0028-3932(87)90120-5

Vasconcelos, R. O., & Ladich, F. (2008). Development of vocalization, auditory sensitivity and acoustic communication in the Lusitanian toadfish Halobatrachus didactylus. *The Journal of Experimental Biology*, *211*(Pt 4), 502–509. https://doi.org/10.1242/jeb.008474

Vergne, A. L., Avril, A., Martin, S., & Mathevon, N. (2007). Parent-offspring communication in the Nile crocodile Crocodylus niloticus: Do newborns' calls show an individual signature? *Die Naturwissenschaften*, *94*(1), 49–54. https://doi.org/10.1007/s00114-006-0156-4

Vergne, A. L., & Mathevon, N. (2008). Crocodile egg sounds signal hatching time. *Current Biology : CB*, *18*(12), R513-4. https://doi.org/10.1016/j.cub.2008.04.011

Vernaleo, B. A., & Dooling, R. J. (2011). Relative salience of envelope and fine structure cues in zebra finch song. *The Journal of the Acoustical Society of America*, *129*(5), 3373–3383. https://doi.org/10.1121/1.3560121

Vernaleo, B. A., Dooling, R. J., & Leek, M. R. (2010). Relative salience of acoustic features in zebra finch song. *The Journal of the Acoustical Society of America*, *127*(3), 1863. https://doi.org/10.1121/1.3384460

Vignal, C., Mathevon, N., & Mottin, S. (2004). Audience drives male songbird response to partner's voice. *Nature*, *430*(6998), 448–451. https://doi.org/10.1038/nature02645

Vignal, C., Mathevon, N., & Mottin, S. (2008). Mate recognition by female zebra finch: Analysis of individuality in male call and first investigations on female decoding process. *Behavioural Processes*, *77*(2), 191–198. https://doi.org/10.1016/j.beproc.2007.09.003

Vitousek, M. N., Adelman, J. S., Gregory, N. C., & Clair, J. J. H. S. (2007). Heterospecific alarm call recognition in a non-vocal reptile. *Biology Letters*, *3*(6), 632–634. https://doi.org/10.1098/rsbl.2007.0443

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

Walcott, C., Evers, D., Froehler, M., & Krakauer, A. (1999). Individuality in “Yodel” Calls Recorded From a Banded Population of Common Loons, Gavia Immer. *Bioacoustics*, *10*(2-3), 101–114. https://doi.org/10.1080/09524622.1999.9753424

Wascher, C. A. F., Hillemann, F., Canestrari, D., & Baglione, V. (2015). Carrion crows learn to discriminate between calls of reliable and unreliable conspecifics. *Animal Cognition*, *18*(5), 1181–1185. https://doi.org/10.1007/s10071-015-0879-8

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

Watanabe, S., Yamamoto, E., & Uozumi, M. (2006). Language discrimination by Java sparrows. *Behavioural Processes*, *73*(1), 114–116. https://doi.org/10.1016/j.beproc.2006.01.013

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, *38*(4), 654–662. https://doi.org/10.1016/j.wocn.2010.10.001

Weishampel, D. B. (1981). Acoustic analyses of potential vocalization in lambeosaurine dinosaurs (Reptilia: Ornithischia). *Paleobiology*, *7*(2), 252–261. https://doi.org/10.1017/S0094837300004036

Weishampel, D. B. (2004). Ornithischia. In D. B. Weishampel, H. Osmlska, & P. Dodson (Eds.), *The Dinosauria*. Berkeley: University of California Press. Retrieved from https://california.universitypressscholarship.com/view/10.1525/california/9780520242098.001.0001/upso-9780520242098

Weishampel, D. B., Osmlska, H., & Dodson, P. (Eds.) (2004). *The Dinosauria*. Berkeley: University of California Press. https://doi.org/10.1525/j.ctt1pn61w

Weissengruber, G. E., Forstenpointner, G., Peters, G., Kübber-Heiss, A., & Fitch, W. T. (2002). Hyoid apparatus and pharynx in the lion (Panthera leo), jaguar (Panthera onca), tiger (Panthera tigris), cheetah (Acinonyxjubatus) and domestic cat (Felis silvestris f. Catus). *Journal of Anatomy*, *201*(3), 195–209. https://doi.org/10.1046/j.1469-7580.2002.00088.x

Wenndt, S. J. (2016). Human recognition of familiar voices. *The Journal of the Acoustical Society of America*, *140*(2), 1172. https://doi.org/10.1121/1.4958682

West, S. A., Griffin, A. S., & Gardner, A. (2006). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*. (20), 415–432.

Wheeler, B. C., Fahy, M., & Tiddi, B. (2019). Experimental evidence for heterospecific alarm signal recognition via associative learning in wild capuchin monkeys. *Animal Cognition*, *22*(5), 687–695. https://doi.org/10.1007/s10071-019-01264-3

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. *Use R*. New York, NY: Springer-Verlag New York. Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10340663 https://doi.org/10.1007/978-0-387-98141-3

Wiley, R. H. (2005). Individuality in songs of Acadian flycatchers and recognition of neighbours. *Animal Behaviour,* *70*(1), 237–247. https://doi.org/10.1016/j.anbehav.2004.09.027

Wilkinson, G. S. (1984). Reciprocal food sharing in the vampire bat. *Nature*, *308*(5955), 181–184. https://doi.org/10.1038/308181a0

Williams, H., Cynx, J., & Nottebohm, F. (1989). Timbre control in zebra finch (Taeniopygia guttata) song syllables. *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *103*(4), 366–380. https://doi.org/10.1037/0735-7036.103.4.366

Witmer, L. M., & Ridgely, R. C. (2008). The paranasal air sinuses of predatory and armored dinosaurs (archosauria: Theropoda and ankylosauria) and their contribution to cephalic structure. *Anatomical Record (Hoboken, N.J. : 2007)*, *291*(11), 1362–1388. https://doi.org/10.1002/ar.20794

Wittenberger, J. F., & Tilson, R. L. (1980). The Evolution of Monogamy: Hypotheses and Evidence. *Annual Review of Ecology and Systematics*, *11*(1), 197–232. https://doi.org/10.1146/annurev.es.11.110180.001213

Wright, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist*. (56), 330–338.

Yang, X.-J., Lei, F., Wang, G., & Jesse, A. J. (2007). Syllable sharing and inter-individual syllable variation in Anna's hummingbird Calypte anna songs, in San Francisco, California. *Folia Zoologica -Praha*. (56), 307–318.

Yao, Y. [Yuan], Lin, Y., Ali, A., & Taylor, C. (2006). Automatic vocal individual recognition of acorn woodpecker (Melanerpesformicivorus) based on hidden Markov models. *Physics Today*, *120*(5), 3000. https://doi.org/10.1121/1.4786991

Yorzinski, J. L., & Vehrencamp, S. L. (2009). The Effect of Predator Type and Danger Level on the Mob Calls of the American Crow. *The Condor*, *111*(1), 159–168. https://doi.org/10.1525/cond.2009.080057

Yosef, R., & Yosef, N. (2010). Cooperative hunting in Brown-Necked Raven (Corvus rufficollis) on Egyptian Mastigure (Uromastyx aegyptius). *Journal of Ethology*, *28*(2), 385–388. https://doi.org/10.1007/s10164-009-0191-7

Yost, W. A. (2000). *Fundamentals of hearing: An introduction* (4. ed.). San Diego: Acad. Press.

Young, B. A. (1991). Morphological basis of "growling" in the king cobra, Ophiophagus hannah. *The Journal of Experimental Zoology*, *260*(3), 275–287. https://doi.org/10.1002/jez.1402600302

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

Yovel, Y., Melcon, M. L., Franz, M. O., Denzinger, A., & Schnitzler, H.-U. (2009). The voice of bats: How greater mouse-eared bats recognize individuals based on their echolocation calls. *PLoS Computational Biology*, *5*(6), e1000400. https://doi.org/10.1371/journal.pcbi.1000400

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Yu, J., Lu, H., Sun, W., Liang, W., Wang, H., & Møller, A. P. (2019). Heterospecific alarm-call recognition in two warbler hosts of common cuckoos. *Animal Cognition*, *22*(6), 1149–1157. https://doi.org/10.1007/s10071-019-01307-9

Yu, X., Peng, Y., Aowphol, A., Ding, L., Brauth, S. E., & Tang, Y.-Z. (2011). Geographic variation in the advertisement calls of Gekko gecko in relation to variations in morphological features: implications for regional population differentiation. *Ethology Ecology & Evolution*, *23*(3), 211–228. https://doi.org/10.1080/03949370.2011.566581

Zann, R. A. (1976). Variation in the Songs of Three Species of Estrildine Grassfinches. *Emu - Austral Ornithology*, *76*(3), 97–108. https://doi.org/10.1071/MU9760097

Zann, R. A. (1996). *The zebra finch: A synthesis of field and laboratory studies*. *Oxford ornithology series: Vol. 5*. Oxford: Oxford Univ. Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy0604/95041163-d.html

Zann, R. A. (1997). Vocal learning in wild and domesticated zebra finches: Signature cues for kin discrimination or epiphenomena? In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development*. Cambridge: Cambridge University Press.

Zsebők, S., Moskát, C., & Bán, M. (2017). Individually distinctive vocalization in Common Cuckoos (Cuculus canorus). *Journal of Ornithology*, *158*(1), 213–222. https://doi.org/10.1007/s10336-016-1376-9

Zug, G. R., Vitt, L. J., & Caldwell, J. P. (2007). *Herpetology: An introductory biology of amphibians and reptiles* (2. ed., repr). San Diego, Calif.: Academic Press.

Appendix A

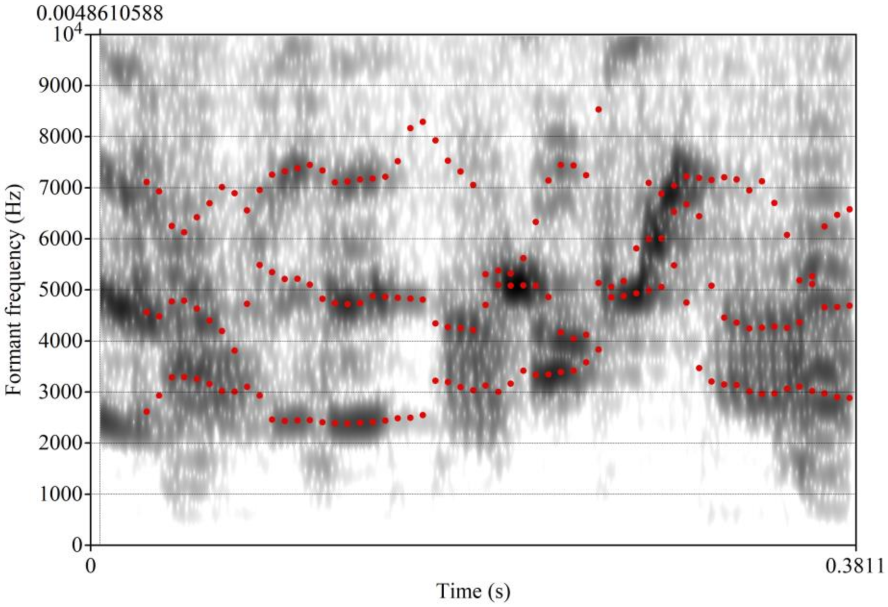


Figure 1: Spectrogram ranging from 0Hz to 10,000Hz showing one motif of zebra finch *A* with extracted formants drawn in (red dots), created in Praat version 6.0.49 (Boersma & Weenink 2019). Red dots overlapping with enhanced areas (see Chapter 1) supports the assumption that Praat has extracted the correct formant frequencies.

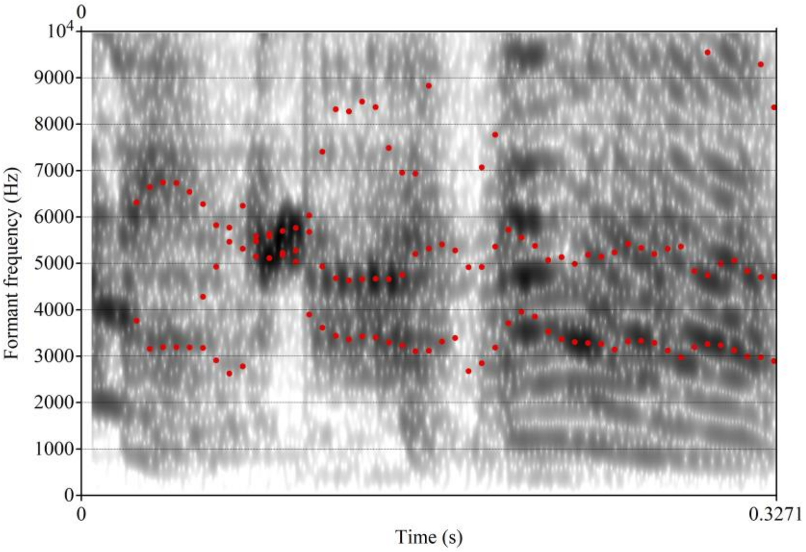


Figure 2: Spectrogram ranging from 0Hz to 10,000Hz showing one motif of zebra finch *B* with extracted formants drawn in (red dots) created in Praat version 6.0.49 (Boersma & Weenink 2019). As only two formants (two rows of red dots) could be extracted, F3 was not further considered in this study.

Table 1: Acoustic features of stimuli from experiment 1.

stimulus	pitch (Hz)	F1 (Hz)	F2 (Hz)	disp. F1-F2 (Hz)
a1	3230	3215	4846	1631
a2	3220	3195	4789	1594
a3	3510	3198	4721	1523
a4	3245	3187	4796	1609
a5	3309	3138	4684	1546
a6	2854	3182	4707	1525
a7	3061	3243	4771	1528
a8	3310	3176	4706	1530
a9	2858	3127	4687	1560
b1	2482	3410	5257	1847
b2	3161	3438	5219	1781
b3	3527	3442	5243	1801
b4	3068	3317	5198	1881
b5	2672	3348	5238	1890
b6	2944	3345	5068	1723
b7	2851	3469	5248	1779
b8	2733	3219	4961	1742
b9	2562	3348	5169	1821

Table 2: Pairs of stimuli from experiment 1 that triggered the responses “false alarm” or “miss” at least three times across all participants and were used to analyse the relevance of the acoustic cues pitch, F1, and formant dispersion.

false alarm	**miss**
b5a8	a2a6
b6a2	a5a3
a3b7	a3a7
a4b5	a4a2
a5b1	a6a5
a5b4	a5a9
a6b9	a4a6
a8b3	a7a9
b3a7	a8a4
b4a7	a4a9
b4a8	b1b5
b5a6	b1b9
b8a2	b1b2
b6a5	b2b2
b1a6	b2b8
b4a6	b3b1
b3a9	b4b4
a4b1	b3b6
a5b2	b7b6
a4b3	b9b6
a9b5	b8b7
a7b6	b8b6
a8b8	

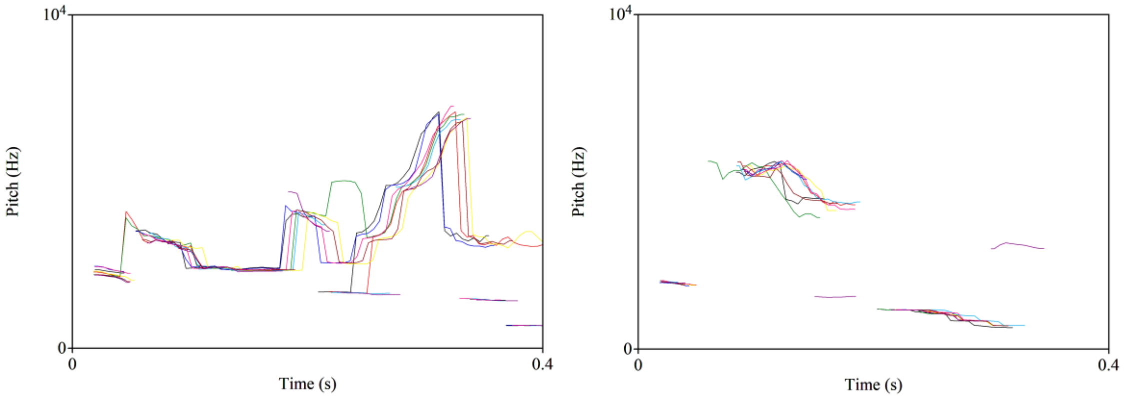


Figure 3: Pitch contour of the stimuli used in experiment 1 produced by zebra finch *A* (left) and zebra finch *B* (right), created in Praat version 6.0.49 (Boersma & Weenink, 2019). Each colour corresponds to one stimulus per zebra finch (black=1, red=2, green=3, blue=4, yellow=5, cyan=6, magenta=7, brown=8, purple=9 as numbered in Appendix A, table 1).

Table 3: Pitch and formant frequencies of stimuli from experiment 2 after removal of pitch contour.

stimulus	**pitch (Hz)**	**F1 (Hz)**	**F2 (Hz)**	**disp. F1-F2 (Hz)**
a1	3168	3233	5914	2681
a2	3152	3214	5648	2434
a3	3427	3457	5053	1596
a4	3188	3311	5552	2241
a5	3264	3301	5324	2023
a6	2815	2930	5427	2497
a7	3030	3090	5603	2513
a8	3255	3269	5578	2309
a9	2817	2907	5561	2654
b1	2531	2747	5122	2375
b2	3048	3194	6032	2838
b3	3571	3442	5611	2169
b4	2969	3075	5887	2812
b5	2783	2847	5513	2666
b6	2939	2917	5628	2711
b7	2865	2924	5622	2698
b8	2582	2790	5380	2590
b9	2565	2729	5273	2544

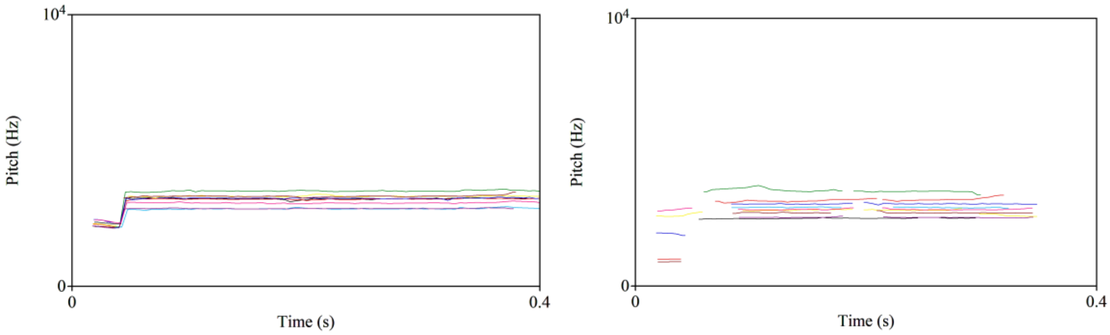


Figure 4: Pitch contour of the modified stimuli used in experiment 2 produced by zebra finch *A* (left) and zebra finch *B* (right) where pitch contour was equalized, created in Praat version 6.0.49 (Boersma & Weenink, 2019). Each colour corresponds to one stimulus per zebra finch (black=1, red=2, green=3, blue=4, yellow=5, cyan=6, magenta=7, brown=8, purple=9 as numbered in Appendix A, table 3).

Table 4: Pairs of stimuli that triggered the same answer (hit, miss, correct reject, false alarm) at least 4 times for incorrect responses ("miss" and "false alarm") and 5 times for correct responses ("hit" and "correct reject") in experiment 2.

pair	hit	miss	reject	false
a4a8	7			
a5a3	9			
b3b3	6			
a5a4	8			
a6a5	7			
a9a2	8	7		
b6b4	8			
a8a7	6			
b3b4	5			
b8b7	5	4		
a2a3	5			
a4a6	5			
a8a2	5			
a8a9	6	4		
b1b3	5			
b4b7	6			
b7b1	5			
b6b3	5			
b1b8	5			

b3b8	8			
b4b8	5			
b5b1	5			
b5b3	5			
b5b8	5			
b6b5	4			
b7b2	4			
a6a8	4			
b1b6	4			
b1b2	4			
b1b4	4			
b6b9	4			
a2b8			7	
a7b3			6	
b2a9			9	
a3b8			5	
b1a4			7	
a4b8			8	
a5b8			7	
a6b2			5	
a7b1			8	
a8b1			5	
a8b6			6	4
b5a3			7	
a8b5			6	
a9b1			5	
b2a4			5	
a2b3			5	
a1b1			6	
a2b4				7
b2a8				5
b3a8				5
b7a9				6
a8b4				4
a9b3				4
b1a3				4

b4a9				4
b7a6				4
a6b8				4
a7b4				4
a3b6				4
a4b7				4

Table 5: Pitch and formant frequencies of stimuli from experiment 3.

stimulus	**pitch (Hz)**	**F1 (Hz)**	**F2 (Hz)**	**disp. F1-F2 (Hz)**
a1	319	1355	2396	1041
a2	384	1306	2225	919
a3	381	1421	2101	680
a4	549	1398	2650	1252
a5	537	1404	2291	887
a6	533	1357	2508	1151
a7	571	1430	2473	1043
a8	508	1399	2453	1054
a9	560	1426	2580	1154
b1	294	1462	3240	1778
b2	382	1408	3223	1815
b3	382	1426	3156	1730
b4	373	1463	2871	1408
b5	385	1486	2993	1507
b6	375	1484	3110	1626
b7	375	1489	2990	1051
b8	383	1473	2825	1352
b9	249	1469	3072	1603

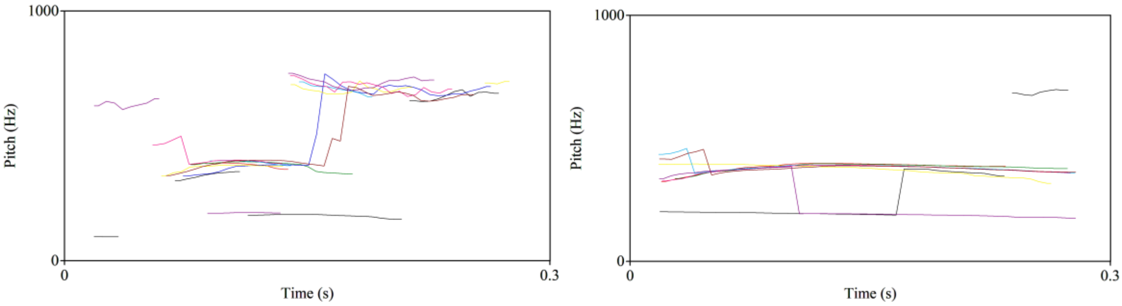


Figure 5: Pitch contour of the stimuli used in experiment 3 produced by crow *A* (left) and crow *B* (right), created in Praat version 6.0.49 (Boersma & Weenink, 2019). Each colour corresponds to one call per crow (black=1, red=2, green=3, blue=4, yellow=5, cyan=6, magenta=7, brown=8, purple=9 as numbered in Appendix A, table 5).

Table 6: Difference in pitch between low pitch and high pitch for crow *A* (excluding the two stimuli that do not contain this pitch transition.

stimuli	**low pitch (Hz)**	**high pitch (Hz)**
a1	181	657
a4	371	683
a5	373	685
a6	394	688
a7	395	696
a8	395	665
a9	191	713

Table 7: Occurrence of each stimulus in pairs that triggered an incorrect response (either “false alarm” or “miss”) out of a total 525 incorrect pairs. Occurrences were counted per pair, so stimuli that appeared twice in one pair (e.g. a1a1) were counted as appearing in one pair instead of occurring twice.

stimulus	**occurrence**
a1	22
a2	37
a3	66
a4	64
a5	60
a6	43
a7	56
a8	54
a9	58
b1	68
b2	85
b3	65
b4	87
b5	58
b6	61
b7	51
b8	74
b9	32

Table 8: Comparison of acoustic parameters between the problematic stimuli b2, b4 and b8 with the mean values for crows *A* and *B*. Frequencies are in Hz, shimmer (local), harmonicity and harmonicity ratios in dB, and jitter (local, absolute) in seconds.

stimulus	**pitch**	**F1**	**disp. F1-F2**	**shimmer**	**jitter**	**harmonicity**	**noise-harm. ratio**	**harm.-noise ratio**
mean b	355	1462	1541	1.14	4.58^{-5}	0.83	0.21	7.93
b2	382	1408	1815	1.03	4.9^{-5}	0.8	0.29	7.32
b4	373	1463	1408	1.38	5.02^{-5}	0.82	0.22	7.22
b8	383	1473	1352	1.37	5.59^{-5}	0.85	0.21	8.54

Table 9: Pairs of stimuli from experiment 4 matched for minimal difference in mean pitch. The pair a1-b9, which had the largest difference in mean pitch, is not included.

a2	b4
a3	b1
a4	b2
a5	b3
a6	b6
a7	b5
a8	b7
a9	b8

Appendix B

Middlesex University London

Psychology REC

The Burroughs
Hendon
London NW4 4BT

Main Switchboard: 0208 411 5000

22/10/2019

APPLICATION NUMBER: 8360

Dear Sabrina Schalz and all collaborators/co-investigators

Re your application title: Heterospecific Discrimination - Quantitative - Evolution

Supervisor: Fabia Tom Dickins Franco

Co-investigators/collaborators:

Thank you for submitting your application. I can confirm that your application has been given APPROVAL from the date of this letter by the Psychology REC.

The following documents have been reviewed and approved as part of this research ethics application:

Document Type	File Name	Date	Version
Parent/Guardian Informed Consent	Parent Consent Form	02/10/2019	1
Participant Information Sheet	Information sheet for adults	02/10/2019	1
Participant Information Sheet	Consent Form adults	02/10/2019	1
Risk Assessment	Risk Assessment Form 5	02/10/2019	1
Permission/Agreement Letter	GateKeeper Letter	02/10/2019	1
Participant Information Sheet	Information sheet for parents	03/10/2019	1
Debriefing Sheet	Debriefing Parents	03/10/2019	1
Debriefing Sheet	Adult debriefing	03/10/2019	1
Participant Recruitment Information	recruitment flyer	03/10/2019	1
Data Protection Act checklist	Data Protection Checklist	03/10/2019	1
GDPR Declaration	GDPR declaration	03/10/2019	1

Although your application has been approved, the reviewers of your application may have made some useful comments on your application. Please look at your online application again to check whether the reviewers have added any comments for you to look at.

Also, please note the following:

1. Please ensure that you contact your supervisor/research ethics committee (REC) if any changes are made to the research project which could affect your ethics approval. There is an Amendment sub-form on MORE that can be completed and submitted to your REC for further review.

2. You must notify your supervisor/REC if there is a breach in data protection management or any issues that arise that may lead to a health and safety concern or conflict of interests.

3. If you require more time to complete your research, i.e., beyond the date specified in your application, please complete the Extension sub-form on MORE and submit it your REC for review.

4. Please quote the application number in any correspondence.

5. It is important that you retain this document as evidence of research ethics approval, as it may be required for submission to external bodies (e.g., NHS, grant awarding bodies) or as part of your research report, dissemination (e.g., journal articles) and data management plan.

6. Also, please forward any other information that would be helpful in enhancing our application form and procedures - please contact

Figure 1: Approval of the ethics application for experiments 1-3 presented in Chapter 1, granted by the Psychology REC 22nd October 2019 prior to data collection.

Middlesex
University
London

Psychology REC

The Burroughs
Hendon
London NW4 4BT

Main Switchboard: 0208 411 5000

27/11/2019

APPLICATION NUMBER: 8360

Dear Sabrina Schaiz and all collaborators/co-investigators

Re your application title: Heterospecific Discrimination - Quantitative - Evolution

Supervisor:

Co-investigators/collaborators: Prof Tom Dickins

Thank you for submitting your application. I can confirm that your application has been given APPROVAL from the date of this letter by the Psychology REC.

The following documents have been reviewed and approved as part of this research ethics application:

Document Type	File Name	Date	Version
Amendments	ethics approval	22/10/2019	1

Although your application has been approved, the reviewers of your application may have made some useful comments on your application. Please look at your online application again to check whether the reviewers have added any comments for you to look at.

Also, please note the following:

1. Please ensure that you contact your supervisor/research ethics committee (REC) if any changes are made to the research project which could affect your ethics approval. There is an Amendment sub-form on MORE that can be completed and submitted to your REC for further review.

2. You must notify your supervisor/REC if there is a breach in data protection management or any issues that arise that may lead to a health and safety concern or conflict of interests.

3. If you require more time to complete your research, i.e., beyond the date specified in your application, please complete the Extension sub-form on MORE and submit it your REC for review.

4. Please quote the application number in any correspondence.

5. It is important that you retain this document as evidence of research ethics approval, as it may be required for submission to external bodies (e.g., NHS, grant awarding bodies) or as part of your research report, dissemination (e.g., journal articles) and data management plan.

6. Also, please forward any other information that would be helpful in enhancing our application form and procedures - please contact MOREsupport@mdx.ac.uk to provide feedback.

Good luck with your research.

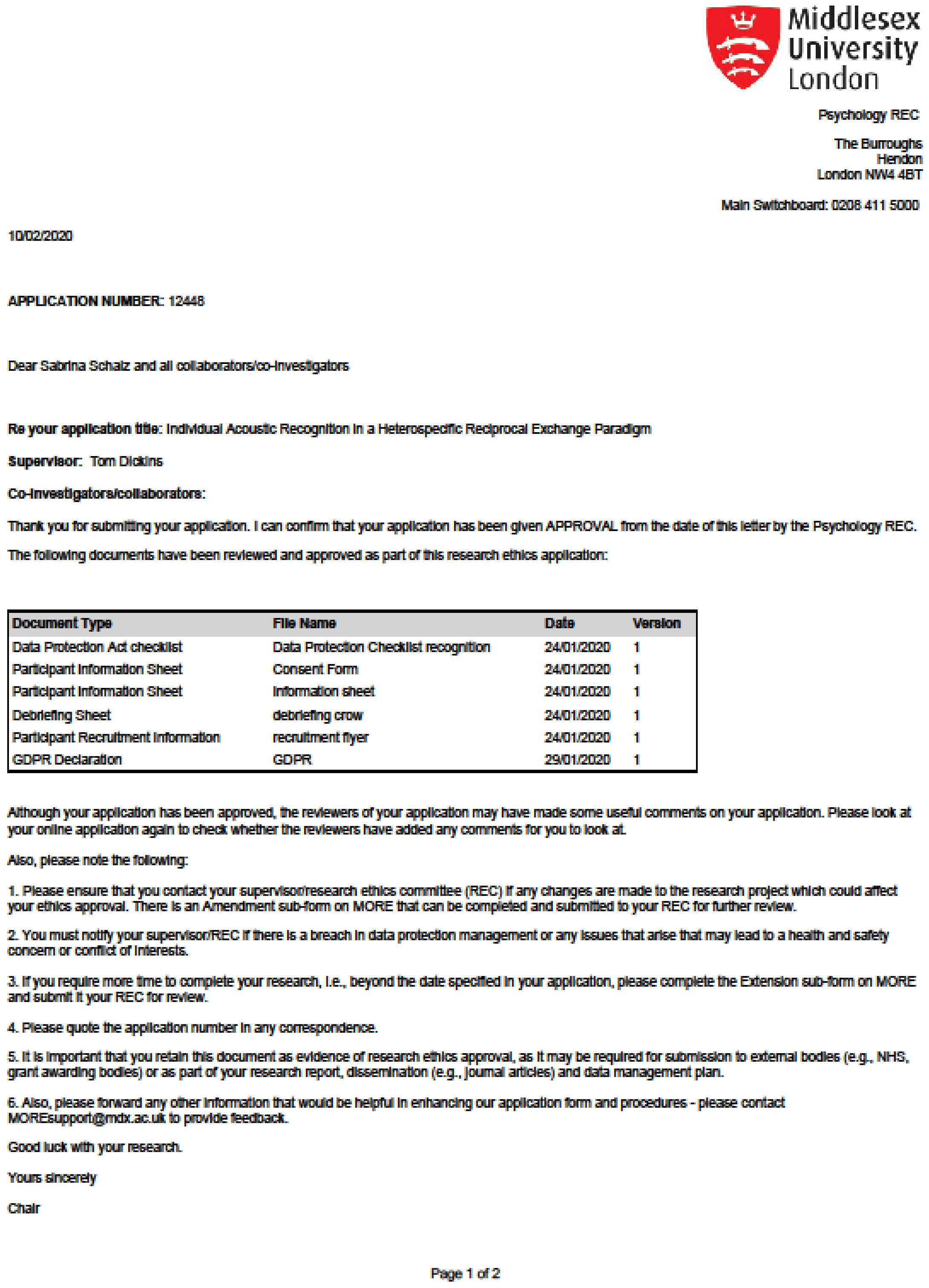
Yours sincerely,

Chair

Psychology REC

Page 1 of 2

Figure 2: Approval of the amendment to the ethics application for experiment 3 presented in Chapter 1, granted by the Psychology REC 27th November 2019 prior to data collection. This application was submitted to change the species whose vocalizations would be used as stimuli (now large-billed crows).



Middlesex
University
London

Psychology REC

The Burroughs
Hendon
London NW4 4BT

Main Switchboard: 0208 411 5000

10/02/2020

APPLICATION NUMBER: 12448

Dear Sabrina Schalz and all collaborators/co-investigators

Re your application title: Individual Acoustic Recognition in a Heterospecific Reciprocal Exchange Paradigm

Supervisor: Tom Dickins

Co-investigators/collaborators:

Thank you for submitting your application. I can confirm that your application has been given APPROVAL from the date of this letter by the Psychology REC.

The following documents have been reviewed and approved as part of this research ethics application:

Document Type	File Name	Date	Version
Data Protection Act checklist	Data Protection Checklist recognition	24/01/2020	1
Participant Information Sheet	Consent Form	24/01/2020	1
Participant Information Sheet	Information sheet	24/01/2020	1
Debriefing Sheet	debriefing crow	24/01/2020	1
Participant Recruitment Information	recruitment flyer	24/01/2020	1
GDPR Declaration	GDPR	29/01/2020	1

Although your application has been approved, the reviewers of your application may have made some useful comments on your application. Please look at your online application again to check whether the reviewers have added any comments for you to look at.

Also, please note the following:

1. Please ensure that you contact your supervisor/research ethics committee (REC) if any changes are made to the research project which could affect your ethics approval. There is an Amendment sub-form on MORE that can be completed and submitted to your REC for further review.

2. You must notify your supervisor/REC if there is a breach in data protection management or any issues that arise that may lead to a health and safety concern or conflict of interests.

3. If you require more time to complete your research, i.e., beyond the date specified in your application, please complete the Extension sub-form on MORE and submit it your REC for review.

4. Please quote the application number in any correspondence.

5. It is important that you retain this document as evidence of research ethics approval, as it may be required for submission to external bodies (e.g., NHS, grant awarding bodies) or as part of your research report, dissemination (e.g., journal articles) and data management plan.

6. Also, please forward any other information that would be helpful in enhancing our application form and procedures - please contact MOREsupport@mdx.ac.uk to provide feedback.

Good luck with your research.

Yours sincerely

Chair

Page 1 of 2

Figure 3: Approval of the ethics application for experiment 4 presented in Chapter 2, granted by the Psychology REC 10th February 2020 prior to data collection.