frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 13 January 2015
doi: 10.3389/fpsyg.2014.01543

Revisiting vocal perception in non-human animals: a review of vowel discrimination, speaker voice recognition, and speaker normalization

Buddhamas Kriengwatana[1,2] *, Paola Escudero[3] and Carel ten Cate[1,2]

[1] *Behavioural Biology, Institute for Biology Leiden, Leiden University, Leiden, Netherlands*
[2] *Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands*
[3] *The MARCS Institute, University of Western Sydney, Sydney, NSW, Australia*

Edited by:
Janet F. Werker, The University of British Columbia, Canada

Reviewed by:
LouAnn Gerken, University of Arizona, USA
Andrew J. Lotto, University of Arizona, USA
Marilyn Vihman, University of York, UK

****Correspondence:***
Buddhamas Kriengwatana, Behavioural Biology, Institute for Biology Leiden, Leiden University, Sylviusweg 72, Leiden 2333BE, Netherlands
e-mail: bkrieng@alumni.uwo.ca

The extent to which human speech perception evolved by taking advantage of predispositions and pre-existing features of vertebrate auditory and cognitive systems remains a central question in the evolution of speech. This paper reviews asymmetries in vowel perception, speaker voice recognition, and speaker normalization in non-human animals – topics that have not been thoroughly discussed in relation to the abilities of non-human animals, but are nonetheless important aspects of vocal perception. Throughout this paper we demonstrate that addressing these issues in non-human animals is relevant and worthwhile because many non-human animals must deal with similar issues in their natural environment. That is, they must also discriminate between similar-sounding vocalizations, determine signaler identity from vocalizations, and resolve signaler-dependent variation in vocalizations from conspecifics. Overall, we find that, although plausible, the current evidence is insufficiently strong to conclude that directional asymmetries in vowel perception are specific to humans, or that non-human animals can use voice characteristics to recognize human individuals. However, we do find some indication that non-human animals can normalize speaker differences. Accordingly, we identify avenues for future research that would greatly improve and advance our understanding of these topics.

Keywords: asymmetries in vowel perception, comparative cognition, general auditory approach, voice perception, language evolution, animal behavior, speaker normalization

INTRODUCTION

The answer to how humans perceive speech has eluded researchers for over half a century (Jusczyk and Luce, 2002; Samuel, 2011). Remarkably, studies in non-human animals (hereafter referred to as animals) have shown that animals can also solve certain problems that are crucial to human speech perception, such as lack of invariance and compensation for co-articulation (e.g., Kluender et al., 1987; Lotto et al., 1997a). The results of these pioneering studies have shown remarkable similarities, but also differences, in perception, discrimination, and sensitivity to acoustic properties of speech in humans and animals (reviewed by Kuhl, 1981; Lotto et al., 1997b; Kluender et al., 2005; Beckers, 2011; Carbonell and Lotto, 2014; ten Cate, 2014). Consequently, many researchers have adopted the general auditory approach outlined by Diehl et al. (2004), which is a framework for the idea that human speech perception is achieved via general learning mechanisms and auditory principles common to humans and animals.

From a general auditory approach, categorical perception occurs at natural psychophysical boundaries constrained by the functioning of the auditory system (Kuhl and Miller, 1975), compensation for coarticulation is possible by contrasting spectral patterns of high and low energy in particular frequency regions (Lotto et al., 1997b; Diehl et al., 2004), and the lack of invariance in speech can be solved in ways similar to concept formation for visual categories that cannot be defined by any single cue (Kluender et al., 1987). Furthermore, phonetic category learning by humans and animals can potentially be achieved via statistical learning and perceptual learning mechanisms. Statistical learning can account for how human infants (Maye et al., 2002), but also rats (Pons, 2006), use the distributional properties of acoustic input to learn phonetic categories, such that exposure to a speech sound continuum with unimodal or bimodal distribution can result in acquisition of one or two phonetic categories, respectively. Statistical learning also appears to underlie our ability to use recurring sound sequences in speech to denote word boundaries (Saffran et al., 1996; Pelucchi et al., 2009), which is also observed when stimuli are tones or musical sounds (Saffran et al., 1999; Gebhart et al., 2009). Thus, statistical learning is a general learning mechanism that is not specific to speech, but is useful for speech perception because it can be used to map acoustic properties onto phonetic categories in a probabilistic manner (Holt et al., 1998).

Altogether, these studies culminate to form the current dominant view that at least several processes involved in speech perception in humans can be traced back to predispositions, learning mechanisms and rudimentary features of vertebrate cognitive and auditory systems also present in other species (e.g., Carbonell and Lotto, 2014). Nevertheless, an enduring and central question in the evolution of speech and language is whether our extraordinary abilities to deal with the enormous variety of speech sound and voices is a matter of degree compared to

the abilities of animals, or the result of an evolutionary quantum leap resulting in novel and unique specialized mechanisms. Animals have, of course, never been under selection to process speech sounds or recognize voices. If they would also process their own communication sounds similar to how humans do, this might indicate a more general mechanism, but might also indicate an independently evolved perceptual mechanism highly specific to their own vocal communication. Hence, a much stronger indication for the presence of general perceptual mechanisms would be if animals process human speech sounds similar to humans. This would indicate that our ability to handle the complexities of speech likely arose from an amalgamation and adaptation of simpler mechanisms present in other vocalizing animals. Comparative studies can thus provide an invaluable window into the uniqueness and origin of speech processing mechanisms.

The objective of this paper is to extend the discussion of species-shared perceptual mechanisms to aspects of human speech perception and voice recognition that have not been considered central to conventional theories of speech perception, but nonetheless cannot be ignored. Specifically, we focus our review on asymmetries in vowel perception, speaker voice recognition, and speaker normalization and underline how these areas of research can benefit from incorporating comparative perspectives. Reviewing these topics contributes significantly to the debate about the nature and specialness (or generality) of human speech perception because asymmetries in vowel perception may play a crucial role in the development of infant speech perception (Polka and Bohn, 2011), voice perception impacts speech perception (e.g., Nygaard and Pisoni, 1998), and intrinsic speaker normalization is just as meaningful an issue as extrinsic speaker normalization because both are concerned with the problem of perceptual invariance in vowel perception (e.g., Johnson, 2005). Thus, our manuscript provides the first detailed review necessary for a more thorough understanding of whether the mechanisms that mediate asymmetries in vowel perception, voice recognition, and speaker normalization in humans arose from an evolutionary quantum leap, or from tuning and remodeling of existing mechanisms that are also present in non-human species (whether by common descent or by independent evolution unconnected to the presence of speech).

ASYMMETRIES IN VOWEL PERCEPTION

Polka and Bohn (2003, 2011) review numerous studies on vowel discrimination in human infants and adults. These studies demonstrate a striking directional asymmetry: discrimination of native or non-native vowels by infants and of non-native vowels by adults is easier when the change is from a vowel occupying a more central position in the F1/F2 vowel space to a vowel occupying a more peripheral position (e.g., from /e/ to /i/). Of the dozens of studies reviewed, only one showed that a change from a more to a less peripheral vowel was easier to discriminate than a change in the reverse direction (Best and Faber, 2000 as cited in Polka and Bohn, 2003, 2011), and this was attributed to effects of vowel rounding in F3. Polka and Bohn (2003, 2011) assert that this directional asymmetry helps infants to acquire phonetic categories because the most peripheral vowels /i/, /a/, and /u/ found in all human languages act as stable referents from which infants can perceptually organize their vowel space (see, Polka and Bohn, 2011 for these ideas in the context of the natural referent vowel framework). The authors propose that these biases are specific to humans due to their role in organizing the vowel space and presence very early in development. Emphatically, Polka and Bohn (2003, 2011) claim that these directional asymmetries do not reflect a general auditory processing bias and therefore will not be present in other animals.

To support their claim of species-specificity in the asymmetries observed in human vowel perception, Polka and Bohn (2003) examined vowel discrimination data from red-winged blackbirds (*Agelaius phoeniceus*), pigeons (*Columba livia*), and cats (Hienz et al., 1981, 1996). While red-winged blackbirds and cats (but not pigeons) also exhibited asymmetries in vowel perception, discrimination was almost always easier when formant frequencies were shifted upward (e.g., from /ɔ/ to /ɑ/ or from /ʊ/ to /æ/).

These results, however, do not resolutely show that the central to peripheral bias found in humans is uniquely human. This is because these experiments do not test discrimination between more and less peripheral vowels. Hienz et al. (1981) tested discrimination of the peripheral vowels /ɔ/, /ɑ/, /æ/, /ɛ/, which does not test whether animals find the change from a central vowel to a peripheral vowel easier to discriminate. Hienz et al. (1996) tested discrimination between the slightly more central vowel /ʊ/ and the peripheral vowels /ɑ/, /æ/, and /ɛ/, and found that the change from /ʊ/ to /ɑ/, /ʊ/ to /æ/, and //ʊ/ to /ɛ/ was easier to discriminate than the change in the opposite direction (e.g., /ɑ/ to /ʊ/). Therefore, these results in fact suggest that animals also find the change from a central vowel to peripheral vowel easier to discriminate. Nevertheless, differences in discrimination by humans and non-human animals on the same speech contrasts are certainly needed to conclude that the asymmetry patterns observed in humans are truly unique to humans. The only contrast that was tested in both humans and red-winged blackbirds was the /ɛ/-/æ/ contrast. In this case, 6–8 and 10–12 month-old human infants found the discrimination easier if the change occurred from /ɛ/ to /æ/, whereas birds found the change in the reverse direction easier to discriminate (Hienz et al., 1981; Polka and Bohn, 1996). This directional asymmetry was recently found in even younger infants (2–3 month-olds) that were exposed to /ɛ/ and /æ/ vowels that were bimodally distributed along an [ɛ-æ] continuum (Wanrooij et al., 2014), and was treated as further evidence to support the central to peripheral bias in humans (Bohn and Polka, 2014). However, as both /ɛ/ and /æ/ occupy peripheral positions in vowel space, this asymmetry can only falsify the central to peripheral bias if we assume that the change from /ɛ/ to /æ/ is easier to discriminate because /æ/ is closer to the vowel referent /a/ than /ɛ/ is to the vowel referent /i/ – according to the natural vowel referent framework, /a/, /i/, and /u/ are vowel referents because they are the most peripheral vowels (Polka and Bohn, 2011). Whether this assumption is correct is not explicitly stated in Polka and Bohn (2003, 2011), but if it is then it is the only vowel contrast so far that could be argued to reflect a human-specific vowel discrimination bias.

Missing from Polka and Bohn's (2003, 2011) reviews was work by Sinnott (1989), who compared detection and discrimination of

synthetic English vowels by human adults and monkeys (vervets *Cercopithecus aethiops* and Japanese macaque *Macaca fuscata*). Subjects had to discriminate between two sets of vowels: /i-ɪ-ɛ-æ-ɜ˞-ʌ-ɑ/ (mostly front vowels with higher F2) and /ʌ-ɑ-ɔ-ʊ-u/ (mostly back vowels with lower F2). Sinnott (1989) used a repeating standard XA task, where subjects pressed a lever to hear a repeating standard vowel followed by two repetitions of a comparison vowel. Subjects had to release the lever if they perceived the comparison vowel to be different from the standard vowel. Every vowel in each of the two sets served as both standard and comparison vowels. We used these data to assess the asymmetries in monkey vowel discrimination (Sinnott, 1989, Table II, p. 560). These directional asymmetries are portrayed visually in **Figure 1**. We only considered directional asymmetries to be present when the difference between the percent of time a vowel comparison was missed following a particular standard was greater than 25. For example, the /u-ʊ/ asymmetry for vervets was included because vervets missed the change from /u/ to /ʊ/ 93 percent of the time, but missed the change from /ʊ/ to /u/ only 2% of the time.

Figure 1 shows that, for back vowels, vervets perform similarly to red-winged blackbirds and cats (Hienz et al., 1981, 1996). That is, their discrimination of back vowels is enhanced if the F1 and F2 of the comparison vowel are increased relative to the standard vowel. However, vervets and macaques (but not adult humans) most likely perceive decreases in F2 of these vowels as decreases in intensity (Sinnott, 1989), and earlier work by Sinnott et al. (1985) showed that vervets and macaques have difficulty detecting decrements in intensity. That is, humans, macaques, and vervets required similar intensities to be able to detect front vowels, but monkeys required back vowels to be presented ∼10–20 dB louder than humans to be able to detect them (Sinnott, 1989). Examination of F1, F2, and F3 values of the vowel stimuli showed that monkeys' detection thresholds were correlated with decreases in F2. This suggested that monkeys required higher intensities in order to be able to detect vowels that decrease in F2 (Sinnott, 1989). Decreased sensitivity to pure tones at lower frequencies of less than 1.0 kHz has also been reported in monkeys compared to humans (Owren et al., 1988). Consequently, perceptual asymmetries involving back vowel contrasts may simply reflect the difficulty that monkeys have if the decrease in F2 from the standard to comparison vowel is perceived as a decrease in intensity. When vowel detection thresholds of humans and monkeys are most similar (i.e., for the front vowels /i/ and /ɪ/; Sinnott, 1989), vervets show the same directional asymmetry as human infants (Swoboda et al., 1978; Dejardins and Trainor, 1998). Therefore, we might expect infants to show a similar directional asymmetry when tested on low back contrasts if they also perceive decreases in F2 as decreases in vowel intensity. This is because infants, like monkeys, are unable to discriminate stimuli that decrease in intensity (Sinnott and Aslin, 1985).

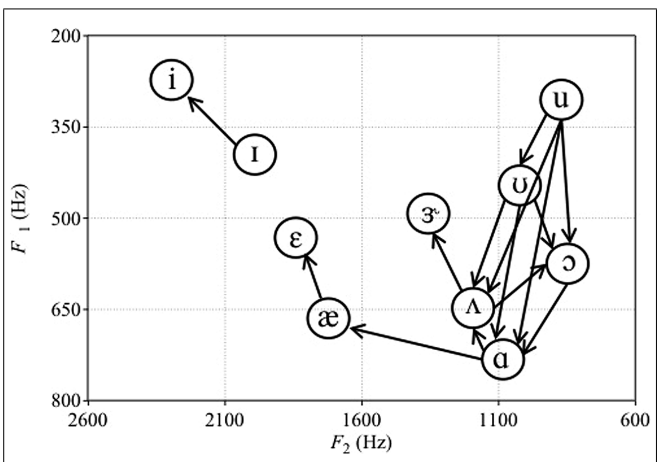

FIGURE 1 | Plot of the asymmetries in vowel discrimination of vervets from Sinnott (1989). Arrows represent the direction of change that was easier to discriminate. For macaques, only the ɑ-ʌ contrast was easier to discriminate than the ʌ-ɑ contrast. F1 and F2 values of vowel tokens are those reported in Sinnott (1989).

Therefore, to demonstrate that the directional asymmetries proposed by Polka and Bohn's (2011) natural referent vowel framework are exclusive to humans, we see an absolute need for more studies on humans and animals that use identical stimuli and comparable experimental designs in order to enable between-species comparisons of asymmetries in vowel perception. In particular, we encourage studies in human infants that test for perceptual asymmetries between low back contrasts that have already been examined in animals, such has the /ʊ/ – /ɑ/ contrast (where cats, red-winged blackbirds, and monkeys find the change from /ʊ/ to /ɑ/ easier than the change from /ɑ/ to /ʊ/), and the /u/ – /ʊ/ and /ɑ/ – /ʌ/ contrasts (where monkeys' performances contradict the predictions of the central-to-peripheral asymmetry hypothesis). Lastly, to properly delineate the function of perceptual asymmetries in vowel perception in humans, we must also understand the causes and possible functions of directional asymmetries in auditory perception in animals. For example, do animals exhibit directional asymmetries in detecting a change in their species-specific communication, and do they make use of these perceptual asymmetries? Or are directional asymmetries a by-product of properties of auditory systems (and thus have no functional use)?

Regardless of whether directional biases in vowel perception are found to be uniquely human, general perceptual biases in human audition may further our understanding of the existence of vowel asymmetries. Various auditory perceptual biases have been found, such as sounds with increasing intensity being perceived as closer than sounds than sounds with decreasing intensity (Neuhoff, 1998), ramped sounds being perceived as having greater intensity and longer duration than damped sounds (Irino and Patterson, 1996; Schlauch et al., 2001), and frequency modulated sounds being easier to detect among pure tones distractors than the reverse (Cusack and Carlyon, 2003). Some of these asymmetries appear to have plausible evolutionary explanations. For instance, rising harmonic sound intensities are reliable indicators of an approaching sound source, thus humans and monkeys perceive the source as being closer than reality when sound intensity increases (but not decreases) in order to account for a "margin of safety" (Neuhoff, 1998; Ghazanfar et al., 2002).

SPEAKER VOICE RECOGNITION

The speech signal not only contains linguistic information, but also nonlinguistic information about the speaker from his/her

voice, such as age, gender, and socio-linguistic background. Voices are sounds generated by vibrations of the vocal folds in the larynx, which are then modified by the vocal tract, leading to an enhancement of particular frequencies (i.e., formant frequencies). The fundamental frequency and formant frequencies in a person's voice are influenced by the size of their vocal folds and vocal tract, which is why men, with their larger vocal folds and vocal tract, tend to have lower fundamental frequencies and formant frequencies than women and children (reviewed in Belin et al., 2004; Latinus and Belin, 2011). Voice differences between speakers may be related to slight variations in the vocal apparatus of different individuals. Differences in the way the vocal apparatus is used, either intentionally or inadvertently, contributes to the distinctiveness of voices. For example, nasal voices are produced when the velum is in a slightly lowered position and breathy voices are produced when the vocal folds remain slightly parted during speaking (see Story and Titze, 2002). Human adults can reliably detect a speaker's approximate age, gender, and accented speech (Hartman, 1979; Flege, 1984; Mullennix et al., 1995; Bachorowski and Owren, 1999), and can also identify familiar speakers from hearing their voices (Bricker and Pruzansky, 1976). Importantly, speaker voice characteristics interact with linguistic comprehension during speech perception: familiar speakers elicit better performance in a word identification task than unfamiliar speakers (Nygaard and Pisoni, 1998), and vowel, word, and consonant-vowel identification are more accurate in single speaker compared to multiple speaker conditions (Strange et al., 1976, 1983; Assmann et al., 1982; Magnuson and Nusbaum, 2007). Knowledge of a speaker's accent and gender also increases word and vowel identification accuracy, respectively (Johnson, 2005; Trude and Brown-Schmidt, 2012; Smith, 2014). Consequently, a complete account of how humans perceive speech also requires consideration of the role that voices play in influencing speech perception.

Perceiving speaker identity from voice information is important for human social interactions. In adults, mechanisms involved in analyzing linguistic and speaker information appear to be partially dissociable (reviewed in Belin et al., 2004, 2011; Belin, 2006). Neuroimaging studies demonstrate that the brain can differentiate "who" from "what" is being said (Formisano et al., 2008), and cortical regions in the mid and anterior superior temporal gyrus respond selectively to human voices and are sensitive to speaker identity (Imaizumi et al., 1997; Belin et al., 2000; Nakamura et al., 2001; Belin and Zatorre, 2003; von Kriegstein et al., 2003). For pre-linguistic infants, extracting the identity of the speaker is likely just as important as deciphering meaning in speech. Even *in utero*, infants respond differently to the voice of their mother compared to a stranger (Kisilevsky et al., 2003), and behavioral and neuroimaging findings confirm that very young infants can indeed discriminate their mother's voice from a stranger's voice (DeCasper and Fifer, 1980; Dehaene-Lambertz et al., 2010), their father's voice from other males (DeCasper and Prescott, 1984), and male from female voices (Jusczyk et al., 1992). Interestingly, the age at which infants become able to identify voices of individuals other than their mother's remains unsettled (Brown, 1979; Hepper et al., 1993; Kisilevsky et al., 2003, 2009; Lee and Kisilevsky, 2014). In any case, these studies indicate that specialization for processing human voices appears to develop over infancy in conjunction with language experience (Grossmann et al., 2010; Vouloumanos et al., 2010; Johnson et al., 2011; Friendly et al., 2014; Schultz et al., 2014).

Humans are not the only species that can extract different types of information from conspecific vocalizations. The ability to distinguish and recognize individual conspecifics in the context of neighbor-stranger recognition has been found in numerous animals (including invertebrates) across the entire animal kingdom (see Tibbetts and Dale, 2007). For instance, Kentucky warblers and hooded warblers know both identity and location of each of their neighbors, and will respond aggressively to neighbor songs that are broadcasted at incorrect territorial boundaries (Godard, 1991; Godard and Wiley, 1995). Female great tits can discriminate between the highly similar songs of their mate and male neighbors (Blumenrath et al., 2007), and may eavesdrop on territorial mate-neighbor singing interactions in order to assess potential extra-pair partners (Otter et al., 1999). Vocal kin recognition has also been repeatedly demonstrated in various birds and mammals (e.g., Rendall et al., 1996; Holekamp et al., 1999; McComb et al., 2000; Illmann et al., 2002; Sharp et al., 2005; Janik et al., 2006; Akçay et al., 2014). Parent-offspring recognition by non-nesting penguins such as the king penguin offers a very strong case of vocal recognition because the chicks must accurately identify the calls of its parents within a crowded and noisy colony with almost no aid from visual or spatial cues. To identify parents, king penguin chicks pay attention to frequency modulations over time in conjunction with a beat analysis (Jouventin et al., 1999; Aubin et al., 2000). In contrast, chicks of nesting penguin species, such as Adelie penguin and gentoo penguin, that can use nesting site as an additional cue for parent identification will use simpler identification mechanisms that rely primarily on pitch and ignore frequency modulations (Jouventin and Aubin, 2002).

The difference between nesting and non-nesting penguins described above highlights the fact that individual recognition by auditory means may be achieved in many ways depending on ecological pressures and evolutionary history. For instance, birds can recognize a conspecific by the songs in his repertoire, by a particular variation in his song, or by his voice characteristics (Weary and Krebs, 1992). European starlings that each have a large song repertoire do not recognize other individual starlings by voice characteristics but rather by memorizing song motifs of different individuals (Gentner and Hulse, 1998; Gentner et al., 2000). On the other hand, great tits also possess a song repertoire but rely on voice characteristics to identify conspecific individuals (Weary and Krebs, 1992). In these experiments, recognition was assessed by training birds to discriminate songs from two individuals and testing their ability to discriminate other previously unheard songs from the same individuals (Weary and Krebs, 1992; Gentner and Hulse, 1998; Gentner et al., 2000). Animals are also capable of heterospecific vocal recognition of familiar individuals, as demonstrated in some monkey species that differentiated between the vocalizations of familiar and unfamiliar members of other monkey species (Candiotti et al., 2013). Observations like this suggests that animals may also be able to distinguish different humans based on their voice characteristics.

Identification of particular human individuals can be useful for animals because they can behave adaptively toward threatening and unthreatening humans by interrupting normal behavior or ignoring them, respectively. Visual discrimination and recognition of individual humans has been documented in several animals such as the magpie (*Pica pica;* Lee et al., 2011), mockingbird (*Mimus polyglottos;* Levey et al., 2009), crow (*Corvus brachyrynchos;* Marzluff et al., 2010, 2012), and octopus (*Enteroctopus dofleini*; Anderson et al., 2010). Some evidence also exists that animals raised in close contact with humans can distinguish between different human voices, which we have summarized in **Table 1**. In most of these studies, recognition was measured by preferential looking times when animals were presented with human faces and a voice that did or did not match one of the faces being shown. The design and interpretation of these studies are similar to those reported for infant studies (see Houston-Price and Nakai, 2004): longer looking times when faces and voices were mismatched were taken as an indication that subject's expectations about the relationship between speaker face and speaker voice (i.e., speaker identity) were violated.

Of the studies in **Table 1**, only Sliwa et al. (2011) and Proops and McComb (2012) have shown that animals can identify human individuals by their voices because the researchers tested discrimination among multiple familiar humans. This is critically important for proving individual identification, as discriminating familiar from unfamiliar individuals is recognition at the class-level, not the individual level (Tibbetts and Dale, 2007). In our opinion, however, even these two studies do not definitively prove human voice recognition because they used the animal's name or phrases that the animal may have been frequently exposed to. This could lead animals to form an association between the specific phrase(s) and the speaker, such that they may not generalize their recognition to novel utterances by the speaker. Thus, animals' performances in these studies may not necessarily reflect genuine recognition of different human speakers based on voice characteristics. We encourage future studies in human voice recognition to incorporate this method of training (i.e., initial discrimination between sounds from two familiar speakers and then testing their ability to recognize different sounds from the same speakers) because it eliminates class-level recognition and learned associations between speaker and specific phrases as confounding factors. Alternatively, researchers may also consider using a habituation-dishabituation paradigm, where subjects are habituated to various speech sounds from one speaker and then tested on whether they dishabituate to speech sounds of a different speaker (e.g., Johnson et al., 2011).

Another fundamental component that these studies do not address is what auditory cues animals may be using to discriminate different human voices, and whether they use the same cues to identify conspecific calls. The question of which cues are used to identify speakers is highly relevant to humans as well, as there is currently no consensus on this matter (see Creel and Bregman, 2011). Some studies have found that adults use pitch and/or formants to identify speakers (e.g., Remez et al., 1987; Fellowes et al., 1997; Baumann and Belin, 2010), whereas others

Table 1 | List of studies that have tested animals' discrimination of different human voices.

Reference	Species	Method	Comparison	Stimulus
Adachi et al. (2007)	Dogs	Face-voice matching	FvU	Animal's name
Sliwa et al. (2011)	Rhesus macaques	Face-voice matching	FvF	Six standardized phrases *(e.g., "bonjour tout le monde," "voila")*
Lampe and Andre (2012)	Horses	Live person-voice matching	FvU	Standardized phrase *"Hey, [animal's name], what are you doing there? Are you having a good day today? We have many riding lessons this week don't we? The semester has started at JMU. You be a good boy/girl today!"*
Proops and McComb (2012)	Horses	Live person-voice matching	FvU FvF	Animal's name
Wascher et al. (2012)	Crows	Playback	FvU	*"Hey"*
Saito and Shinozuka (2013)	Cats	Habituation–dishabituation	FvU	Animal's name
Ratcliffe et al. (2014)	Dogs	Live person-voice matching	Male vs. female	Four standardized phrases *"Hey!," "Come on then," "Good dog!," "What's this?"*
McComb et al. (2014)	Elephants	Playback	Male vs. female Man vs. boy Masaai vs. Kamba	standardized sentence *"Look, look over there, a group of elephants is coming"*

**Note: FvU, familiar versus unfamiliar voice, FvF, familiar versus familiar voice.*

have suggested that these spectral cues are important for gender determination while temporal patterns are used for individual identification (Fellowes et al., 1997). Furthermore, not all listeners use the same cues to distinguish voices, and different cues may be more important for distinguishing particular voices (Kreiman et al., 1992).

We are not aware of any studies that investigate what acoustic cues animals may be using to discriminate or identify human voices, although this is surely a question that merits rigorous investigation. Work by Zoloth et al. (1979) suggests that conspecifics and heterospecifics may attend to different cues, as Japanese macaques used peak of a frequency inflection while other monkey species used initial pitch to discriminate between coos of other Japanese macaques. But if animals use the same cues for recognition of conspecific vocalization as heterospecific vocalization, then we expect that only some species will use voice characteristics for individual recognition. That is, following the example described above, we would expect great tits but not European starlings to recognize different human voices because great tits discriminate conspecifics based on voice characteristics while starlings discriminate conspecifics based on song motifs (Weary and Krebs, 1992; Gentner and Hulse, 1998; Gentner et al., 2000). Even then, great tits may rely on different parameters than human listeners, as the song parameters that show significant variation between individual great tits are the number of phrases in a song, duration of the first phase minus the last phase, maximum frequency, and pitch (Weary, 1990; Weary et al., 1990). The acoustic similarities between human speech and species-specific vocalizations may also determine which cues are used. For example, zebra finches distance calls have harmonic structures that resemble vowel formants in human speech (Dooling et al., 1995). This similarity may explain why zebra finches perform similarly to humans during discrimination of speech sounds (Dooling et al., 1995; Ohms et al., 2012), and why neurons in a secondary auditory area of the zebra finch brain respond more strongly to human speech and species-specific calls than calls of other songbirds that are acoustically less similar (Chew et al., 1996).

Findings from animal studies are valuable for understanding the extent to which human voices can be differentiated and recognized without the need for human voice-specific neural networks (Leaver and Rauschecker, 2010; Belin et al., 2011). However, research on human voice processing in animals to date is scarce, and of the studies that do exist, the ability to recognize voices (defined as discriminating between different familiar voices, as opposed to discriminating between familiar and unfamiliar voices) has yet to be compellingly demonstrated. An understanding of the differences and similarities of acoustic cues and mechanisms used for acoustic recognition of conspecifics and heterospecifics is also currently severely lacking. We strongly believe that these issues must be addressed in future studies if we are to discover parallels between voice processing in humans and animals, and consequently shed light onto the evolution of human voice perception.

SPEAKER NORMALIZATION

The counterpart of being able to extract information about speaker identity is being able to handle acoustic variations of the same utterance caused by speaker differences. The acoustic realizations of phonemes and words can vary tremendously between speakers, due to physical, contextual, environmental, and sociolinguistic factors (i.e., age and gender differences in vocal tract size and shape, coarticulation, background noise, and accents). Consequently, speaker normalization refers to our ability to recognize phonologically identical utterances despite high acoustic variability across speakers (Johnson, 2005). A compelling example of the immense variability in the speech signal resulting from differences between speakers is in vowel production. Vowels are reliably distinguished by the first and second formant frequencies (F1 and F2); however, F1 and F2 values of vowels produced by different speakers (and especially different genders) are highly variable within a vowel category and greatly overlap between categories, to the extent that the acoustic distance within a vowel category can be just as large as the acoustic distance between vowel categories (Potter and Steinberg, 1950; Peterson and Barney, 1952; Hillenbrand et al., 1995). Consequently, studies that seek to understand our impressive ability to normalize vowels despite this intensive overlap (Peterson and Barney, 1952; Strange et al., 1976, 1983; Assmann et al., 1982) can provide convincing evidence of what processes contribute to speaker normalization.

Researchers have not yet reached a consensus on how vowel normalization in humans is achieved (reviewed in Nearey, 1989; Johnson et al., 1999; Adank, 2003; Johnson, 2005). Some argue that normalization occurs via low-level auditory perceptual processes, by computation of particular formant ratios that allow vowel categories to be distinctively represented in discrete regions in acoustic space (Potter and Steinberg, 1950; Syrdal and Gopal, 1986; Miller, 1989), or by using F0 or F3 values (i.e., fundamental frequency and third formant frequency) that are correlated with vocal tract length to disambiguate ambiguous F1 and F2 values (e.g., Ladefoged and Broadbent, 1957; Fujisaki and Kawashima, 1968; Wakita, 1977; Nearey, 1989). A recent paper suggests that humans normalize for speaker differences by computing ratios between F1 and F2 in relation to F3. Specifically, Monahan and Idsardi (2010) showed that transforming F1 and F2 values into F1/F3 and F2/F3 ratios effectively eliminated variation between speakers in a corpus of American English vowels from Hillenbrand et al. (1995). They also provided neurophysiological evidence that the human brain is sensitive to F1/F3 ratios, complementing prior work showing that the brain is sensitive to F2/F3 ratios (Monahan and Idsardi, 2010). Alternatively, others argue that listeners form rich perceptual representations of speaker identity by incorporating vocal tract length with other learned factors such as familiarity or socio-cultural expectations; these abstract speaker representations subsequently influence vowel normalization (i.e., "talker normalization"; Johnson, 1990; Johnson et al., 1999).

Both types of views have merits and drawbacks, and they have also both received empirical support. The auditory perceptual approaches can account for how vowel normalization occurs with limited linguistic capabilities and familiarity with novel speakers (Monahan and Idsardi, 2010). Behavioral and neurophysiological studies indicate that humans normalize speaker differences in vowels without linguistic comprehension or attention. In adults, extraction and processing of vowel formants takes place at a

subcortical and pre-attentional level (von Kriegstein et al., 2006; Monahan and Idsardi, 2010; Tuomainen et al., 2013), and that even pre-linguistic infants can categorize vowels of different speakers and genders (Kuhl, 1983). On the other hand, talker normalization approaches explain how listeners can learn speaker-specific and language-specific patterns of speech (Johnson et al., 1999). Several findings support the idea that learning of non-acoustic speaker-related variables is indeed involved in accommodating for speaker differences in phonetic realizations (Samuel and Kraljic, 2009). For instance, expectations of speaker gender can alter phoneme boundaries (Johnson et al., 1999), single speaker and familiar speakers conditions yield better vowel identification (Strange et al., 1976, 1983; Assmann et al., 1982; Mullennix et al., 1989; Nygaard and Pisoni, 1998), and infants and adults can rapidly adapt to accented speech (Clarke and Garrett, 2004; Bradlow and Bent, 2008; White and Aslin, 2011; Cristia et al., 2012; van Heugten and Johnson, 2014)

Comparative studies in animals represent one way to address whether speaker normalization mechanisms are unique to humans. Although normalization is necessary for word recognition, normalization also occurs for phonetic segments. Indeed, normalization has been demonstrated in 6 month-old infants before they acquire words, and at the level of vowels (Kuhl, 1983). That is, non-verbal infants can normalize speech even before they have learned words. Even for the few words that 6–9 month-old infants do seem to recognize (Bergelson and Swingley, 2012), normalization may not have been required for recognition, since the words were spoken by a familiar speaker (i.e., a parent). Therefore, it is important to discern what perceptual processes infants use to normalize speech, and whether these mechanisms are rudimentary processes that are available to infants because they are species-shared or non-speech related perceptual mechanisms. Unfortunately, only a handful of animal studies have considered speaker variability as a factor, even though the ability to account for speaker-dependent variation is fundamental for speech perception. A possible reason for lack of interest in how animals handle speaker-dependent variation in speech may have to do with skepticism regarding what, if any, benefits this investigation would yield. In other words, as suggested by Trout (2001), what reason would an animal have to normalize speaker variation in speech? What advantages would it confer to the animal? And even if animals do appear to normalize speech, are they applying similar mechanisms to humans? We argue that studying speaker normalization in non-human animals is ecologically valid because animals also have to deal with signal variability when recognizing vocalizations made by other individuals of the same species (conspecifics) and other individuals of a different species (heterospecifics). On the other hand, whether animals and humans attend to the same cues and have shared mechanisms for normalizing human speech is an open and exciting question that is yet to be answered.

Signal variability in animal vocalizations can be caused by inter-individual differences, such as physical size (Fitch, 1997, 1999; Growcott et al., 2011), intra-individual differences, such as affective state (e.g., stress; Perez et al., 2012), and a combination of inter-and intra-individual differences such as fluctuations in endocrine state (Galeotti et al., 1997; Soma et al., 2002). Yet animals must still be able to acquire information about the type of vocalization (such as calls indicating predator versus food source) and in some contexts, the identity of the signaler. Thus, in parallel to how humans can extract linguistic information despite speaker-dependent variation in the speech signal, animals can also categorize vocalizations of conspecific as well as heterospecific vocalizations despite individual variation in the signal. Indeed, responding to hetereospecific signals is widespread in many animals because of the advantages it confers to the receiver (Seppänen et al., 2007). For example, red-breasted nuthatches (*Sitta canadensis*) are sensitive to variations in black-capped chickadee (*Poecile atricapillus*) mobbing calls that encode information about the size and degree of threat of predators (Templeton and Greene, 2007), avian brood parasites may eavesdrop on sexual signals of host species to assess parental quality (Parejo and Avilés, 2007), and hornbills can respond to differentially to alarm calls of Diana monkeys that signal significant and non-significant threats for the hornbills (Rainey et al., 2004).

The red-breasted nuthatches' extraction of information from the black-capped chickadee *chick-a-dee* call is particularly compelling because of the highly sophisticated nature of this vocalization, both at a contextual and an acoustic level. In addition to signaling predators, the *chick-a-dee* call is used in other contexts, such as to maintain group cohesion and coordinate group movements (Ficken et al., 1978). Significant individual differences in the *chick-a-dee* call can occur in every 100-Hz interval between 500 and 7000 Hz, with greater variation between individuals than within individuals of the same group in various temporal and spectral parameters (Mammen and Nowicki, 1981). Acoustically, the call is made of four note types (A, B, C, D) sung in a fixed sequence, but note types can be repeated or omitted to create 100s of variations such as ACCDD, AABBCD, or ADDDD (Hailman et al., 1985). Black-capped chickadees can also modify spectral components of the D note so that this note converges amongst members of the same group *(*Nowicki, 1989*)*.

Importantly, the same variations in syntax (number of D notes), temporal and spectral characteristics (duration and interval between D notes, frequency overtone spacing and bandwidth) are also used to convey information about predator threat (Templeton et al., 2005). That is, the same acoustic properties that encode predator threat also vary depending on the individual. Yet despite the apparent overlap between acoustic parameters that signal individual/group identity and predator threat, nuthatches are still able to obtain information relevant for their behaviors. Thus, this constitutes a naturally occurring example of normalization of heterospecific vocal signals by a non-human animal (i.e., disregarding irrelevant variation caused by individual differences in the vocalizations of another species). In this light, the idea that non-human animals can account for speaker differences in human vocalizations seems quite plausible, and in the following sections we review studies that suggest that speaker normalization of speech is not a uniquely human ability.

The ability to normalize speaker differences in naturally spoken vowels or correlates of gender differences in synthetic vowels (i.e., F0, which is generally higher in female than male voices and is utilized in perception of voice gender by humans; Hillenbrand et al., 1995; Mullennix et al., 1995; Lavner et al., 2000) has been tested

in cats, rhesus monkeys, dogs, chimpanzees, chinchillas, budgerigars, rats, and ferrets (Dewson, 1964; Dewson et al., 1969; Baru, 1975; Burdick and Miller, 1975; Kojima and Kiritani, 1989; Dooling and Brown, 1990; Eriksson and Villa, 2006; Bizley et al., 2013). All of these studies reported that animals were able to differentiate stimuli based on vowel category and ignore speaker-dependent variation. However, we argue that none of these studies robustly demonstrate speaker normalization. This is because the experiments on cats, dogs, rhesus monkeys, chimpanzees, chinchillas, and budgerigars tested discrimination of the vowels /i/, /u/, and or /a/ (Dewson, 1964; Dewson et al., 1969; Baru, 1975; Burdick and Miller, 1975; Kojima and Kiritani, 1989). As these vowels occupy distant acoustic spaces, the variability between these vowel categories is likely larger than the variability between speakers. Consequently, a stronger test for vowel normalization would use vowel categories where speaker differences and vowel category differences have relatively equal variation. This was addressed by Dooling and Brown (1990), who found that budgerigars could discriminate between the psychoacoustically closer vowels /i/ and /ɛ/ produced by different male speakers. However, they did not examine whether budgerigars could also discriminate these vowels produced by different female speakers, so we cannot be sure that budgerigars can in fact normalize these vowels because much of the acoustic overlap between vowel categories are due to gender differences (see Peterson and Barney, 1952). Lastly, the experiments in rats, ferrets, and chimpanzees used synthetic vowels that differed only in one acoustic parameter (Kojima and Kiritani, 1989; Eriksson and Villa, 2006; Bizley et al., 2013). This is problematic for conclusions about speaker normalization because voices differ in various dimensions, and no single acoustic cue has been found that reliably predicts speaker characteristics that can be used to normalize vowels.

As far as we know, Ohms et al. (2010) are the only researchers that have clearly demonstrated that an animal can normalize vowels of different speakers. Specifically, Ohms et al. (2010) found that zebra finches can generalize their discrimination of two words that differ only in vowels (wɪt and wɛt) to multiple, different male or female speakers after being trained to discriminate these words from a single speaker of the same sex. Their results show that birds are indeed learning something about the phonemic differences between wɪt and wɛt, and are able to apply these criteria to unfamiliar speakers while ignoring speaker-dependent differences in production of these words. Notably, zebra finches could also recognize the same word produced by male and female speakers after learning the word from a set of speakers of the other sex (Ohms et al., 2010; see also ten Cate, 2014). Remarkably, this ability seems lacking in human infants (Houston and Jusczyk, 2000). A logical follow up of these results would be a test for normalization of isolated vowels – an ability demonstrated by human adults (Strange et al., 1976, 1983; Assmann et al., 1982). This is because the birds in Ohms et al. (2010) could have been using other acoustic cues instead of vowel differences during generalization, such as formant transitions between consonant and vowel. Recently, a study by Engineer et al. (2013) reported that rats too could normalize speaker variation in consonants. Rats could learn to discriminate between the words *dad* and *tad* produced by a female speaker and generalize this discrimination to novel male and female speakers. With the same stimuli, rats could also learn to distinguish between the word *dad* spoken by a female speaker and the same word that was pitch-shifted down by one octave, and subsequently generalize this discrimination to novel male and female voices (Engineer et al., 2013). This shows that, like humans, animals can extract different types of information from speech (see section on speaker voice recognition).

Griebel and Oller (2012) provide another interesting case that may potentially reflect normalization of speaker differences by a Yorkshire terrier (Bailey). They found that Bailey could correctly retrieve 13 out of 16 familiar toys when verbally requested by a female experimenter with a German accent and a male experimenter with a western American English (California) accent, even though Bailey's owner was female with a southern American English (Tennessee) accent. Bailey's accurate performance with these accented voices was not due to training or familiarity, as the experimenters had never previously requested the toys from Bailey. Again, we do not know which acoustic parameters Bailey was using to recognize familiar words spoken by unfamiliar speakers. The authors note that toy names often consisted of "two or more words that included intonation and alliteration or assonance cues that may have made them easier to remember and discriminate" (Griebel and Oller, 2012). This suggests that Bailey may not have needed to normalize speaker differences in vowel production in order to recognize the words, but possibly relied on prosodic cues – another feature in speech that tamarins (Ramus et al., 2000), rats (Toro et al., 2003), Java sparrows (Naoi et al., 2012), and zebra finches (Spierings and ten Cate, 2014) are sensitive to.

Other studies on word discrimination in dogs have been conducted, but do not offer clear evidence for speaker normalization. This is because these studies do not control for non-verbal cues from the trainer that dogs could rely on to perform correct actions (see Mills, 2005), or they test dogs' ability to retrieve objects from the verbal commands of a single familiar trainer (Warden and Warner, 1928; Kaminski et al., 2004; Fukuzawa et al., 2005). Another study did not explicitly state in the methods whether commands were consistently given by a single trainer or by multiple trainers (Ramos and Ades, 2012). In a similar vein, Warfield et al. (1966) showed that cats could discriminate the words "bat" and "cat," but they did not test whether discrimination was affected if word tokens were produced by multiple speakers.

Responding to categories of heterospecific communication signals is commonly found in the animal kingdom, and though few studies provide direct evidence, there is some indication that animals may be able to normalize speaker differences in human speech. Yet replication and extension of positive findings such as those by Ohms et al. (2010) and Engineer et al. (2013) are compulsory. In particular, we point out that the ability to categorize speech sounds from different speakers and disregard non-essential information is not an absolute demonstration of normalization, which is why researchers must test whether animals and humans apply the same normalization mechanisms. For example, do both humans and animals successfully categorize vowels of multiple speakers by computing formant ratios as proposed by Monahan and Idsardi (2010), or by using other normalization algorithms that have been previously proposed (reviewed in Escudero and

Bion, 2007). Such tests would be invaluable in verifying whether similar behaviors exhibited by humans and animals are mediated by the same or different mechanisms.

CONCLUSION

In this review we have discussed the current state of animal research in three aspects of speech and voice perception. We have presented arguments and evidence that caution against premature conclusions about whether asymmetries in vowel perception reflect an innate and uniquely human bias that is present in inexperienced language learners and not attributable to general properties of the vertebrate or mammalian auditory system. We have noted that there is a lack of definitive evidence for animal recognition of individual human voices in the literature, and we have suggested ways to improve experimental designs that will more assuredly test whether animals can use voice characteristics to discern different humans. Lastly, we infer with reservation from two recent experiments that animals may be able to normalize speaker differences. Consequently, we strongly encourage researchers to conduct more carefully designed and rigorously controlled experiments to validate the human-specific claims of asymmetries in vowel perception, voice perception, and speaker normalization that we have described. Studies that identify what acoustic cues animals rely on to perform either individual voice perception or speaker normalization are also seriously needed as they are invaluable for our understanding of how these behaviors are accomplished.

With accumulating empirical results, we can sooner reach the stage where findings from these three areas can be synthesized to tackle broader questions in speech perception. An example would be whether and at what level of processing speaker identification and speaker normalization mechanisms interact during speech perception. Some researchers believe that in humans the analysis of linguistic information and speaker identity during speech perception may be segregated into dissociable but interacting neural pathways (although the stage at which the integration of these two streams occurs is undetermined; Belin et al., 2004). Comparative research on whether animals also analyze conspecific vocalizations and/or human speech for communicative content and signaler identity separately would reveal whether or not this compartmentalization occurred as a distinctive human adaptation that enables us to map overlapping and highly variable acoustic information onto correct phonetic categories while simultaneously processing speaker identity related cues.

We have emphasized throughout this paper that addressing these topics in animals is neither insignificant nor extraneous because many social animals encounter similar challenges to those humans face when discriminating similar-sounding vocalizations/phonemes, determining signaler/speaker characteristics from vocalizations/speech, and resolving between-individual variation in order to perceive the content of vocalizations/speech. It is our hope that this review will have cogently demonstrated that expanding our view to include how animals perceive speech can offer valuable insights for more thorough conceptualization of the specificity, simplicity (or complexity), and specialization of human speech perception mechanisms.

ACKNOWLEDGMENT

We would like to thank Gabriël Beckers for comments on an earlier version of this manuscript. This research was supported by Australian Research Council grant DP130102181 (CI Paola Escudero).

REFERENCES

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

Adank, P. (2003). *Vowel Normalization: a Perceptual-Acoustic Study of Dutch Vowels.* Ph.D. thesis, University of Nijmegen, Nijmegen, The Netherlands.

Akçay, Ç., Hambury, K. L., Arnold, J. A., Nevins, A. M., and Dickinson, J. L. (2014). Song sharing with neighbours and relatives in a cooperatively breeding songbird. *Anim. Behav.* 92, 55–62. doi: 10.1016/j.anbehav.2014.03.029

Anderson, R. C., Mather, J. A., Monette, M. Q., and Zimsen, S. R. M. (2010). Octopuses (*Enteroctopus dofleini*) recognize individual humans. *J. Appl. Anim. Welf. Sci.* 13, 261–272. doi: 10.1080/10888705.2010.483892

Assmann, P. F., Nearey, T. M., and Hogan, J. T. (1982). Vowel identification: orthographic, perceptual, and acoustic aspects. *J. Acoust. Soc. Am.* 71, 975–989. doi: 10.1121/1.387579

Aubin, T., Jouventin, P., and Hildebrand, C. (2000). Penguins use the two-voice system to recognize each other. *Proc. R. Soc. Lond.* 267, 1081–1087. doi: 10.1098/rspb.2000.1112

Bachorowski, J. A., and Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *J. Acoust. Soc. Am.* 106, 1054–1063. doi: 10.1121/1.427115

Baru, A. V. (1975). "Discrimination of synthesized vowels [a] and [i] with varying parameters (Fundamental frequency, intensity, duration and number of formants) in dog," in *Auditory Analysis and Perception of Speech*, eds G. Fant and M. A. A. Tatham (Waltham, MA: Academic Press), 91–101.

Baumann, O., and Belin, P. (2010). Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychol. Res.* 74, 110–120. doi: 10.1007/s00426-008-0185-z

Beckers, G. J. L. (2011). Bird speech perception and vocal production: a comparison with humans. *Hum. Biol.* 83, 191–212. doi: 10.3378/027.083.0204

Belin, P. (2006). Voice processing in human and non-human primates. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 2091–2107. doi: 10.1098/rstb.2006.1933

Belin, P., Bestelmeyer, P. E. G., Latinus, M., and Watson, R. (2011). Understanding voice perception. *Br. J. Psychol.* 102, 711–725. doi: 10.1111/j.2044-8295.2011.02041.x

Belin, P., Fecteau, S., and Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* 8, 129–35. doi: 10.1016/j.tics.2004.01.008

Belin, P., and Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal-lobe. *Neuroreport* 14, 2105–2109. doi: 10.1097/00001756-200311140-00019

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–12. doi: 10.1038/35002078

Bergelson, E., and Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3253–3258. doi: 10.1073/pnas.1113380109

Best, C. T., and Faber, A. (2000). "Developmental increase in infants' discrimination of nonnative vowels that adults assimilate to a single native vowel," in *Paper presented at the International Conference on Infant Studies*, Brighton, 16–19.

Bizley, J. K., Walker, K. M. M., King, A. J., and Schnupp, J. W. H. (2013). Spectral timbre perception in ferrets: discrimination of artificial vowels under different listening conditions. *J. Acoust. Soc. Am.* 133, 365–376. doi: 10.1121/1.4768798

Blumenrath, S. H., Dabelsteen, T., and Pedersen, S. B. (2007). Vocal neighbour–mate discrimination in female great tits despite high song similarity. *Anim. Behav.* 73, 789–796. doi: 10.1016/j.anbehav.2006.07.011

Bohn, O.-S., and Polka, L. (2014). Fast phonetic learning in very young infants: what it shows and what it doesn't show. *Front. Psychol.* 5:511. doi: 10.3389/fpsyg.2014.00511

Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* 106, 707–729. doi: 10.1016/j.cognition.2007.04.005

Bricker, P. D., and Pruzansky, S. (1976). "Speaker recognition," in *Contemporary Issues in Experimental Phonetics*, ed. N. J. Lass (New York: Academic), 295–326.

Brown, C. J. (1979). Reactions of infants to their parents' voices. *Infant Behav. Dev.* 2, 295–300. doi: 10.1016/S0163-6383(79)80038-7

Burdick, C. K., and Miller, J. D. (1975). Speech perception by the chinchilla: discrimination of sustained /a/ and /i/. *J. Acoust. Soc. Am.* 58, 415–427. doi: 10.1121/1.380686

Candiotti, A., Zuberbühler, K., and Lemasson, A. (2013). Voice discrimination in four primates. *Behav. Process.* 99, 67–72. doi: 10.1016/j.beproc.2013.06.010

Carbonell, K. M., and Lotto, A. J. (2014). Speech is not special... again. *Front. Psychol.* 5:427. doi: 10.3389/fpsyg.2014.00427

Chew, S. J., Vicario, D. S., and Nottebohm, F. (1996). A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc. Natl. Acad. Sci. U.S.A.* 93, 1950–1955. doi: 10.1073/pnas.93.5.1950

Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* 116, 3647–3658. doi: 10.1121/1.1815131

Creel, S. C., and Bregman, M. R. (2011). How Talker Identity Relates to Language Processing. *Lang. Lingist. Compass* 5, 190–204. doi: 10.1111/j.1749-818X.2011.00276.x

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., and Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Front. Psychol.* 3:479. doi: 10.3389/fpsyg.2012.00479

Cusack, R., and Carlyon, R. P. (2003). Perceptual asymetries in audition. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 713–725. doi: 10.1037/0096-1523.29.3.713

DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: newborns prefer their mother's voices. *Science* 208, 1174–1176. doi: 10.1126/science.7375928

DeCasper, A. J., and Prescott, P. A. (1984). Human newborns' perception of male voices: preference, discrimination, and reinforcing value. *Dev. Psychobiol.* 17, 481–91. doi: 10.1002/dev.420170506

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., et al. (2010). Language or music, mother or Mozart? structural and environmental influences on infants' language networks. *Brain Lang.* 114, 53–65. doi: 10.1016/j.bandl.2009.09.003

Dejardins, R. N., and Trainor, L. J. (1998). Fundamental frequency influences vowel discrimination in 6-month-old infants. *Can. Acoust. Proc.* 26, 96–97.

Dewson, J. H. (1964). Speech sound discrimination by cats. *Science* 144, 555–556. doi: 10.1126/science.144.3618.555

Dewson, J. H., Pribram, K. H., and Lynch, J. C. (1969). Effects of ablations of temporal cortex upon speech sound discrimination in the monkey. *Exp. Neurol.* 24, 579–591. doi: 10.1016/0014-4886(69)90159-9

Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). Speech perception. *Annu. Rev. Psychol.* 55, 149–79. doi: 10.1146/annurev.psych.55.090902.142028

Dooling, R. J., Best, C. T., and Brown, S. D. (1995). Discrimination of synthetic full-formant and sinewave /ra-la/ continua by *Budgerigars* (Melopsittacus undulatus) and *Zebra finches* (Taeniopygia guttata). *J. Acoust. Soc. Am.* 97, 1839–1846. doi: 10.1121/1.412058

Dooling, R. J., and Brown, S. D. (1990). Speech perception by budgerigars (*Melopsittacus undulatus*): spoken vowels. *Percept. Psychophys.* 47, 568–574. doi: 10.3758/BF03203109

Engineer, C. T., Perez, C. A., Carraway, R. S., Chang, K. Q., Roland, J. L., Sloan, A. M., and Kilgard, M. P. (2013). Similarity of cortical activity patterns predicts generalization behavior. *PLoS ONE* 8:e78607. doi: 10.1371/journal.pone.0078607

Eriksson, J. L., and Villa, A. E. P. (2006). Learning of auditory equivalence classes for vowels by rats. *Behav. Process.* 73, 348–59. doi: 10.1016/j.beproc.2006.08.005

Escudero, P., and Bion, R. (2007). "Modeling vowel normalization and sound perception as sequential processes," in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 1413–1416.

Fellowes, J. M., Remez, R. E., and Rubin, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Percept. Psychophys.* 59, 839–849. doi: 10.3758/BF03205502

Ficken, M. S., Ficken, R. W., and Witkin, S. R. (1978). Vocal repertoire of the black-capped chickadee. *Auk* 95, 34–48.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* 102, 1213–1222. doi: 10.1121/1.421048

Fitch, W. T. (1999). Acoustic exaggeration of size in birds by tracheal elongation: comparative and theoretical analyses. *J. Zool.* 248, 31–49. doi: 10.1111/j.1469-7998.1999.tb01020.x

Flege, J. M. (1984). The detection of French accent by American listeners. *J. Acoust. Soc. Am.* 76, 692–707. doi: 10.1121/1.391256

Formisano, E., De Martino, F., Bonte, M., and Goebel, R. (2008). "Who" is saying "What"? Brain-based decoding of human voice and speech. *Science* 322, 970–973. doi: 10.1126/science.1164318

Friendly, R. H., Rendall, D., and Trainor, L. J. (2014). Learning to differentiate individuals by their voices: infants' individuation of native- and foreign-species voices. *Dev. Psychobiol.* 56, 228–37. doi: 10.1002/dev.21164

Fujisaki, H., and Kawashima, T. (1968). The roles of pitch and higher formants in the perception of vowels. *IEEE T. Speech.* 16, 73–77. doi: 10.1109/TAU.1968.1161952

Fukuzawa, M., Mills, D. S., and Cooper, J. J. (2005). The effect of human command phonetic characteristics on auditory cognition in dogs (*Canis familiaris*). *J. Comp. Psychol.* 119, 117–120. doi: 10.1037/0735-7036.119.1.117

Galeotti, P., Saino, N., Sacchi, R., and Møller, A. P. (1997). Song correlates with social context, testosterone and body condition in male barn swallows. *Anim. Behav.* 53, 687–700. doi: 10.1006/anbe.1996.0304

Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychon. Bull. Rev.* 16, 486–490. doi: 10.3758/PBR.16.3.486

Gentner, T. Q., and Hulse, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Anim. Behav.* 56, 579–594. doi: 10.1006/anbe.1998.0810

Gentner, T. Q., Hulse, S. H., Bentley, G. E., and Ball, G. F. (2000). Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. *J. Neurobiol.* 42, 117–133. doi: 10.1002/(SICI)1097-4695(200001)42:1<117::AID-NEU11>3.0.CO;2-M

Ghazanfar, A. A., Neuhoff, J. G., and Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15755–15757. doi: 10.1073/pnas.242469699

Godard, R. (1991). Long-term memory of individual neighbours in a migratory songbird. *Nature* 350, 228–229. doi: 10.1038/350228a0

Godard, R., and Wiley, R. H. (1995). Individual recognition of song repertoires in two wood warblers. *Behav. Ecol. Sociobiol.* 3, 119–123. doi: 10.1007/BF00164157

Griebel, U., and Oller, D. K. (2012). Vocabulary learning in a Yorkshire terrier: slow mapping of spoken words. *PLoS ONE* 7:e30182. doi: 10.1371/journal.pone.0030182

Grossmann, T., Oberecker, R., Koch, S. P., and Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron* 65, 852–858. doi: 10.1016/j.neuron.2010.03.001

Growcott, A., Miller, B., Sirguey, P., Slooten, E., and Dawson, S. (2011). Measuring body length of male sperm whales from their clicks: the relationship between inter-pulse intervals and photogrammetrically measured lengths. *J. Acoust. Soc. Am.* 130, 68–73. doi: 10.1121/1.3578455

Hailman, J. P., Ficken, M. S., and Ficken, R. W. (1985). The 'chick-a-dee' calls of *Parus atricapillus*: a recombinant system of animal communication compared with written english. *Semiotica* 56, 191–224. doi: 10.1515/semi.1985.56.3-4.191

Hartman, D. E. (1979). The perceptual identity and characteristics of aging in normal male adult speakers. *J. Commun. Disord.* 12, 53–61. doi: 10.1016/0021-9924(79)90021-2

Hepper, P., Scott, D., and Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *J. Reprod. Infant Psychol.* 11, 147–153. doi: 10.1080/02646839308403210

Hienz, R. D., Aleszczyk, C. M., and May, B. J. (1996). Vowel discrimination in cats: acquisition, effects of stimulus level, and performance in noise. *J. Acoust. Soc. Am.* 99, 3656–3668. doi: 10.1121/1.414980

Hienz, R. D., Sachs, M. B., and Sinnott, J. M. (1981). Discrimination of steady-state vowels by blackbirds and pigeons. *J. Acoust. Soc. Am.* 70, 699–706. doi: 10.1121/1.386933

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American english vowels. *J. Acoust. Soc. Am.* 97, 3099–3111. doi: 10.1121/1.411872

Holekamp, K., Boydston, E., Szykman, M., Graham, I., Nutt, K., Birch, S., et al. (1999). Vocal recognition in the spotted hyena and its possible implications regarding the evolution of intelligence. *Anim. Behav.* 58, 383–395. doi: 10.1006/anbe.1999.1157

Holt, L. L., Lotto, A. J., and Kluender, K. R. (1998). Incorporating principles of general learning in theories of language acquisition. *Chicago Linguist. Soc.* 34, 253–268.

Houston, D. M., and Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1570–1582. doi: 10.1037/0096-1523.26.5.1570

Houston-Price, C., and Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant Child Dev.* 13, 341–348. doi: 10.1002/icd.364

Illmann, G., Schrader, L., Spinka, M., and Sustr, P. (2002). Acoustical mother-offspring recognition in pigs (*Sus scrofa domestica*). *Behaviour* 139, 487–505. doi: 10.1163/15685390260135970

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport* 8, 2809–2812. doi: 10.1097/00001756-199708180-00031

Irino, T., and Patterson, R. D. (1996). Temporal asymmetry in the auditory system. *J. Acoust. Soc. Am.* 99, 2316–2331. doi: 10.1121/1.415419

Janik, V. M., Sayigh, L. S., and Wells, R. S. (2006). Signature whistle shape conveys identity information to bottlenose dolphins. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8293–8297. doi: 10.1073/pnas.0509918103

Johnson, E. K., Westrek, E., Nazzi, T., and Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Dev. Sci.* 14, 1002–1011. doi: 10.1111/j.1467-7687.2011.01052.x

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *J. Acoust. Soc. Am.* 88, 642–654. doi: 10.1121/1.399767

Johnson, K. (2005). "Speaker normalization in speech perception," in *Handbook of Speech Perception*, eds D. B. Pisoni and R. E. Remez (Oxford: Blackwell), 363–389. doi: 10.1002/9780470757024.ch15

Johnson, K., Strand, E. A., and D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *J. Phon.* 27, 359–384. doi: 10.1006/jpho.1999.0100

Jouventin, P., and Aubin, T. (2002). Acoustic systems are adapted to breeding ecologies: individual recognition in nesting penguins. *Anim. Behav.* 64, 747–757. doi: 10.1006/anbe.2002.4002

Jouventin, P., Aubin, T., and Lengagne, T. (1999). Finding a parent in a king penguin colony: the acoustic system of individual recognition. *Anim. Behav.* 57, 1175–1183. doi: 10.1006/anbe.1999.1086

Jusczyk, P. W., and Luce, P. A. (2002). Speech perception and spoken word recognition: past and present. *Ear Hear.* 23, 2–40. doi: 10.1097/00003446-200202000-00002

Jusczyk, P. W., Pisoni, D. B., and Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition* 43, 253–291. doi: 10.1016/0010-0277(92)90014-9

Kaminski, J., Call, J., and Fischer, J. (2004). Word learning in a domestic dog: evidence for "fast mapping". *Science* 304, 1682–1683. doi: 10.1126/science.1097859

Kisilevsky, B. S., Hains, S. M. J., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., et al. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behav. Dev.* 32, 59–71. doi: 10.1016/j.infbeh.2008.10.002

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychol. Sci.* 14, 220–224. doi: 10.1111/1467-9280.02435

Kluender, K. R., Diehl, R. L., and Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science* 237, 1195–1197. doi: 10.1126/science.3629235

Kluender, K. R., Lotto, A. J., and Holt, L. L. (2005). "Contributions of nonhuman animal models to understanding human speech perception," in *Listening to Speech: An Auditory Perspective*, eds S. Greenberg and W. Ainsworth (New York, NY: Oxford University Press), 203–220.

Kojima, S., and Kiritani, S. (1989). Vocal-auditory functions in the chimpanzee: vowel perception. *Int. J. Primatol.* 10, 199–213. doi: 10.1007/BF02735200

Kreiman, J., Gerratt, B. R., Precoda, K., and Berke, G. S. (1992). Individual differences in voice quality perception. *J. Speech Hear. Res.* 35, 512–520. doi: 10.1044/jshr.3503.512

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound categories. *J. Acoust. Soc. Am.* 70, 340–349. doi: 10.1121/1.386782

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behav. Dev.* 6, 263–285. doi: 10.1016/S0163-6383(83)80036-8

Kuhl, P. K., and Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* 190, 69–72. doi: 10.1126/science.1166301

Ladefoged, P., and Broadbent, D. (1957). Information conveyed by vowels. *J. Acoust. Soc. Am.* 29, 98–104. doi: 10.1121/1.1908694

Lampe, J. F., and Andre, J. (2012). Cross-modal recognition of human individuals in domestic horses (*Equus caballus*). *Anim. Cogn.* 15, 623–630. doi: 10.1007/s10071-012-0490-1

Latinus, M., and Belin, P. (2011). Human voice perception. *Curr. Biol.* 21, R143–R145. doi: 10.1016/j.cub.2010.12.033

Lavner, Y., Gath, I., and Rosenhouse, J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Commun.* 30, 9–26. doi: 10.1016/S0167-6393(99)00028-X

Leaver, A. M., and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612. doi: 10.1523/jneurosci.0296-10.2010

Lee, G. Y., and Kisilevsky, B. S. (2014). Fetuses respond to father's voice but prefer mother's voice after birth. *Dev. Psychobiol.* 56, 1–11. doi: 10.1002/dev.21084

Lee, W. Y., Lee, S., Choe, J. C., and Jablonski, P. G. (2011). Wild birds recognize individual humans: experiments on magpies, Pica pica. *Anim. Cogn.* 14, 817–825. doi: 10.1007/s10071-011-0415-4

Levey, D. J., London, G. A., Poulsen, J. R., Stracey, C. M., and Robinson, S. K. (2009). Urban mockingbirds quickly learn to identify. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8959–8962. doi: 10.1073/pnas.0811422106

Lotto, A. J., Kluender, K. R., and Holt, L. L. (1997a). Perceptual Compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *J. Acoust. Soc. Am.* 102, 1134–1140. doi: 10.1121/1.419865

Lotto, A. J., Kluender, K. R., and Holt, L. L. (1997b). Animal models of speech perception phenomena. *Chicago Linguist. Soc.* 33, 357–367.

Magnuson, J. S., and Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 391–409. doi: 10.1037/0096-1523.33.2.391

Mammen, D. L., and Nowicki, S. (1981). Individual differences and within-flock convergence in *Chickadee Calls. Behav. Ecol. Sociobiol.* 9, 179–186. doi: 10.1007/BF00302935

Marzluff, J. M., Miyaoka, R., Minoshima, S., and Cross, D. J. (2012). Brain imaging reveals neuronal circuitry underlying the crow's perception of human faces. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15912–15917. doi: 10.1073/pnas.1206109109

Marzluff, J. M., Walls, J., Cornell, H. N., Withey, J. C., and Craig, D. P. (2010). Lasting recognition of threatening people by wild American crows. *Anim. Behav.* 79, 699–707. doi: 10.1016/j.anbehav.2009.12.022

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3

McComb, K., Moss, C., Sayialel, S., and Baker, L. (2000). Unusually extensive networks of vocal recognition in African elephants. *Anim. Behav.* 59, 1103–1109. doi: 10.1006/anbe.2000.1406

McComb, K., Shannon, G., Sayialel, K. N., and Moss, C. (2014). Elephants can determine ethnicity, gender, and age from acoustic cues in human voices. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5433–5438. doi: 10.1073/pnas.1321543111

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *J. Acoust. Soc. Am.* 85, 2114–2134. doi: 10.1121/1.397862

Mills, D. S. (2005). What's in a word? a review of the attributes of a command affecting the performance of pet dogs. *Anthrozoös* 18, 208–221. doi: 10.2752/089279305785594108

Monahan, P. J., and Idsardi, W. J. (2010). Auditory Sensitivity to Formant Ratios: Toward an Account of Vowel Normalization. *Lang. Cogn. Process.* 25, 808–839. doi: 10.1080/01690965.2010.490047

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., and Farnsworth, L. M. (1995). The perceptual representation of voice gender. *J. Acoust. Soc. Am.* 98, 3080–3095. doi: 10. 1121/1.413832

Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *J. Acoust. Soc. Am.* 85, 365–378. doi: 10.1121/1.397688

Nakamura, K., Kawashima, R., Sugiura, M., Nakamura, A., Hatano, K., Nagumo, S., et al. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia* 39, 1047–1054. doi: 10.1016/S0028-3932(01)00037-9

Naoi, N., Watanabe, S., Maekawa, K., and Hibiya, J. (2012). Prosody discrimination by songbirds (*Padda oryzivora*). *PLoS ONE* 7:e47446. doi: 10.1371/journal.pone.0047446

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *J. Acoust. Soc. Am.* 85, 2088–2113. doi: 10.1121/1.397861

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature* 395, 123–124. doi: 10.1038/25862

Nowicki, S. (1989). Vocal plasticity in captive black-capped chickadees: the acoustic basis and rate of call convergence. *Anim. Behav.* 38, 64–73. doi: 10.1016/0003-3472(89)90007-9

Nygaard, L. C., and Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Percept. Psychophys.* 60, 355–376. doi: 10.3758/BF03206860

Ohms, V. R., Escudero, P., Lammers, K., and ten Cate, C. (2012). Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception. *Anim. Cogn.* 15, 155–61. doi: 10.1007/s10071-011-0441-2

Ohms, V. R., Gill, A., Van Heijningen, C. A. A., Beckers, G. J. L., and ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proc. R. Soc. Lond. B Biol. Sci.* 277, 1003–1009. doi: 10.1098/rspb.2009.1788

Otter, K., Mcgregor, P. K., Terry, A. M. R., Burford, F. R. L., Peake, T. M., and Dabelsteen, T. (1999). Do female great tits (Parus major) assess males by eavesdropping? A field study using interactive song playback. *Proc. R. Soc. Lond. B Biol. Sci.* 266, 1305–1309. doi: 10.1098/rspb.1999.0779

Owren, M. J., Hopp, S. L., Sinnott, J. M., and Petersen, M. R. (1988). Absolute auditory thresholds in three old world monkey species (*Cercopithecus aethiops*, *C. neglectus, Macaca fuscata*) and humans (*Homo sapiens*). *J. Comp. Psychol.* 102, 99–107. doi: 10.1037/0735-7036.102.2.99

Parejo, D., and Avilés, J. M. (2007). Do avian brood parasites eavesdrop on heterospecific sexual signals revealing host quality? a review of the evidence. *Anim. Cogn.* 10, 81–88. doi: 10.1007/s10071-006-0055-2

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Dev.* 80, 674–85. doi: 10.1111/j.1467-8624.2009.01290.x

Perez, E. C., Elie, J. E., Soulage, C. O., Soula, H. A., Mathevon, N., and Vignal, C. (2012). The acoustic expression of stress in a songbird: does corticosterone drive isolation-induced modifications of zebra finch calls? *Horm. Behav.* 61, 573–581. doi: 10.1016/j.yhbeh.2012.02.004

Peterson, G. E., and Barney, H. H. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175–184. doi: 10.1121/1.1906875

Polka, L., and Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *J. Acoust. Soc. Am.* 100, 577–592. doi: 10.1121/1.415884

Polka, L., and Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Commun.* 41, 221–231. doi: 10.1016/S0167-6393(02)00105-X

Polka, L., and Bohn, O.-S. (2011). Natural referent vowel (NRV) framework: an emerging view of early phonetic development. *J. Phon.* 39, 467–478. doi: 10.1016/j.wocn.2010.08.007

Pons, F. (2006). The effects of distributional learning on rats' sensitivity to phonetic information. *J. Exp. Psychol. Anim. Behav. Process.* 32, 97–101. doi: 10.1037/0097-7403.32.1.97

Potter, R. K., and Steinberg, J. C. (1950). Toward the specification of speech. *J. Acoust. Soc. Am.* 22, 807–820. doi: 10.1121/1.1906694

Proops, L., and McComb, K. (2012). Cross-modal individual recognition in domestic horses (*Equus caballus*) extends to familiar humans. *Proc. R. Soc. Lond. B Biol. Sci.* 279, 3131–3138. doi: 10.1098/rspb.2012.0626

Rainey, H. J., Zuberbühler, K., and Slater, P. J. B. (2004). Hornbills can distinguish between primate alarm calls. *Proc. R. Soc. Lond. B Biol. Sci.* 271, 755–759. doi: 10.1098/rspb.2003.2619

Ramos, D., and Ades, C. (2012). Two-item sentence comprehension by a dog (*Canis familiaris*). *PLoS ONE* 7:e29689. doi: 10.1371/journal.pone.0029689

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

Ratcliffe, V. F., McComb, K., and Reby, D. (2014). Cross-modal discrimination of human gender by domestic dogs. *Anim. Behav.* 91, 127–135. doi: 10.1016/j.anbehav.2014.03.009

Remez, R. E., Rubin, P. E., Nygaard, L. C., and Howell, W. A. (1987). Perceptual normalization of vowels produced by sinusoidal voices. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 40–61. doi: 10.1037/0096-1523.13.1.40

Rendall, D., Rodman, P. S., and Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim. Behav.* 51, 1007–1015. doi: 10.1006/anbe.1996.0103

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52. doi: 10.1016/S0010-0277(98)00075-4

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

Samuel, A. G. (2011). Speech perception. *Annu. Rev. Psychol.* 62, 49–72. doi: 10.1146/annurev.psych.121208.131643

Samuel, A. G., and Kraljic, T. (2009). Perceptual learning for speech. *Atten. Percept. Psychophys.* 71, 1207–1218. doi: 10.3758/APP.71.6.1207

Schlauch, R. S., Ries, D. T., and DiGiovanni, J. J. (2001). Duration discrimination and subjective duration for ramped and damped sounds. *J. Acoust. Soc. Am.* 109, 2880–2887. doi: 10.1121/1.1372913

Schultz, S., Vouloumanos, A., Bennett, R. H., and Pelphrey, K. (2014). Neural specialization for speech in the first months of life. *Dev. Sci.* 17, 766–774. doi: 10.1111/desc.12151

Seppänen, J. T., Forsman, J. T., Mönkkönen, M., and Thomson, R. L. (2007). Social information use is a process across time, space, and ecology, reaching heterospecifics. *Ecology* 88, 1622–1633. doi: 10.1890/06-1757.1

Sharp, S. P., McGowan, A., Wood, M. J., and Hatchwell, B. J. (2005). Learned kin recognition cues in a social bird. *Nature* 434, 1127–30. doi: 10.1038/nature03522

Sinnott, J. M. (1989). Detection and discrimination of synthetic English vowels by Old World monkeys (*Cercopithecus*, Macaca) and humans. *J. Soc. Acoust. Am.* 86, 557–565. doi: 10.1121/1.398235

Sinnott, J. M., and Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *J. Acoust. Soc. Am.* 78, 1986–1992. doi: 10.1121/1.392655

Sinnott, J. M., Petersen, R., and Hopp, S. L. (1985). Frequency and intensity discrimination in humans and monkeys. *J. Acoust. Soc. Am.* 78, 1977–1985. doi: 10.1121/1.392654

Sliwa, J., Duhamel, J.-R., Pascalis, O., and Wirth, S. (2011). Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1735–1740. doi: 10.1073/pnas.1008169108

Smith, D. R. R. (2014). Does knowing speaker sex facilitate vowel recognition at short durations? *Acta Psychol.* 148, 81–90. doi: 10.1016/j.actpsy.2014.01.010

Soma, K. K., Wissman, A. M., Brenowitz, E. A., and Wingfield, J. C. (2002). Dehydroepiandrosterone (DHEA) increases territorial song and the size of an associated brain region in a male songbird. *Horm. Behav.* 41, 203–212. doi: 10.1006/hbeh.2001.1750

Spierings, M. J., and ten Cate, C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proc. R. Soc. Lond. B Biol. Sci.* 281:20140480. doi: 10.1098/rspb.2014.0480

Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *J. Acoust. Soc. Am.* 74, 695–705. doi: 10.1121/1.389855

Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (1976). Consonant environment specifies vowel identity. *J. Acoust. Soc. Am.* 60, 213–224. doi: 10.1121/1.381066

Story, B. H., and Titze, I. R. (2002). A preliminary study of voice quality transformation based on modifications to the neutral vocal tract area function. *J. Phon.* 30, 485–509. doi: 10.1006/jpho.2002.0168

Swoboda, P. J., Kass, J., Morse, P. A., and Leavitt, L. A. (1978). Memory factors in infant vowel discrimination of normal and at-risk infants. *Child Dev.* 49, 332–339. doi: 10.2307/1128695

Syrdal, A. K., and Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *J. Acoust. Soc. Am.* 79, 1086–1100. doi: 10.1121/1.393381

Templeton, C. N., and Greene, E. (2007). Nuthatches eavesdrop on variations in heterospecific chickadee mobbing alarm calls. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5479–5482. doi: 10.1073/pnas.0605183104

Templeton, C. N., Greene, E., and Davis, K. (2005). Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science* 308, 1934–1937. doi: 10.1126/science.1108841

ten Cate, C. (2014). On the phonetic and syntactic processing abilities of birds: From songs to speech and artificial grammars. *Curr. Opin. Neurobiol.* 28, 157–164 doi: 10.1016/j.conb.2014.07.019

Tibbetts, E. A., and Dale, J. (2007). Individual recognition: it is good to be different. *Trends Ecol. Evol.* 22, 529–37. doi: 10.1016/j.tree.2007.09.001

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

Trout, J. D. (2001). The biological basis of speech: what to infer from talking to the animals. *Psychol. Rev.* 108, 523–549. doi: 10.1037/0033-295X.108.3.523

Trude, A. M., and Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Lang. Cogn. Process.* 27, 979–1001. doi: 10.1080/01690965.2011.597153

Tuomainen, J., Savela, J., Obleser, J., and Aaltonen, O. (2013). Attention modulates the use of spectral attributes in vowel discrimination: behavioral and event-related potential evidence. *Brain Res.* 1490, 170–183. doi: 10.1016/j.brainres.2012.10.067

van Heugten, M., and Johnson, E. K. (2014). Learning to contend with accents in infancy: benefits of brief speaker exposure. *J. Exp. Psychol. Gen.* 143, 340–350. doi: 10.1037/a0032192

von Kriegstein, K., Eger, E., Kleinschmidt, A., and Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res. Cogn. Brain Res.* 17, 48–55. doi: 10.1016/S0926-6410(03)00079-X

von Kriegstein, K., Warren, J. D., Ives, D. T., Patterson, R. D., and Griffiths, T. D. (2006). Processing the acoustic effect of size in speech sounds. *NeuroImage* 32, 368–375. doi: 10.1016/j.neuroimage.2006.02.045

Vouloumanos, A., Hauser, M. D., Werker, J. F., and Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Dev.* 81, 517–27. doi: 10.1111/j.1467-8624.2009.01412.x

Wakita, H. (1977). Normalization of vowels by vocal-tract length and its application to vowel identification. *IEEE Trans. Acoustic Speech. Signal Process* 25, 183–192. doi: 10.1109/TASSP.1977.1162929

Wanrooij, K., Boersma, P., and van Zuijen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. *Front. Psychol.* 5:77. doi: 10.3389/fpsyg.2014.00077

Warden, C. J., and Warner, L. H. (1928). The sensory capacities and intelligence of dogs, with a report on the ability of the noted dog "Fellow" to respond to verbal stimuli. *Q. Rev. Biol.* 3, 1–28. doi: 10.1086/394292

Warfield, D., Ruben, R. J., and Glackin, R. (1966). Word discrimination in cats. *J. Aud. Res.* 6, 97–120.

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

Weary, D. M. (1990). Categorization of song notes in great tits: which acoustic features are used and why? *Anim. Behav.* 39, 450–457. doi: 10.1016/S0003-3472(05)80408-7

Weary, D., Falls, J., and McGregor, P. (1990). Song matching and the perception of song types in great tits, *Parus major*. *Behav. Ecol.* 1, 43–47. doi: 10.1093/beheco/1.1.43

Weary, D. M., and Krebs, J. R. (1992). Great tits classify songs by individual voice characteristics. *Anim. Behav.* 43, 283–287. doi: 10.1016/S0003-3472(05)80223-4

White, K. S., and Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Dev. Sci.* 14, 372–384. doi: 10.1111/j.1467-7687.2010.00986.x

Zoloth, S. R., Petersen, M. R., Beecher, M. D., Green, S., Marler, P., Moody, D. B., et al. (1979). Species-specific perceptual processing of vocal Sounds. *Science* 204, 870–873. doi: 10.1126/science.108805

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 September 2014; accepted: 12 December 2014; published online: 13 January 2015.

Citation: Kriengwatana B, Escudero P and ten Cate C (2015) Revisiting vocal perception in non-human animals: a review of vowel discrimination, speaker voice recognition, and speaker normalization. Front. Psychol. ***5****:1543. doi: 10.3389/fpsyg.2014.01543*

This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology.