

Investigating voice identity discrimination in the dog brain using fast periodic auditory stimulation

OTDK work

written by:

DOROTTYA SZILVIA RÁCZ

Biology masters

Neuroscience and Human Biology specialization

Eötvös Loránd University

supervisors:

DR. ATTILA ANDICS

senior research fellow

Department of Ethology

DR. BOGLÁRKA MORVAI

research fellow

Department of Ethology

EÖTVÖS LORÁND UNIVERSITY

FACULTY OF SCIENCE

INSTITUTE OF BIOLOGY



Budapest, 2024

Table of Contents

Abstract.....	3
Introduction.....	3
Conspecific voice (identity) processing among species	3
The effect of conspecificity and experiential relevance on voice (identity) processing.....	5
Voice identity processing in dogs, humans and within the dog-human relationship	6
Fast periodic auditory stimulation for investigating voice identity discrimination.....	8
Materials and methods	10
Participants.....	10
Stimuli	11
Experimental design and stimulus presentation	13
EEG acquisition	15
Data analysis.....	16
<i>Preprocessing</i>	16
<i>Frequency domain analysis</i>	16
<i>Statistical analysis</i>	17
Results	18
Dogs	18
Humans	20
Discussion.....	24
Conclusion	31
References	32
Acknowledgements	40

Abstract

Voices of conspecifics, typically the most relevant social partners, are processed preferentially in many species' brains. However, it is unclear whether this sensitivity reflects an ancient neural tuning to conspecific sounds, or emerges in response to experientially relevant voices. To disentangle these accounts, the special interspecific relationship of dogs and humans provides an ideal study case: for dogs, that are not only selected to communicate effectively with humans but also live with them, human vocalizations may have gained greater social relevance than anciently relevant conspecific vocalizations; and among humans, dog vocalizations may be more relevant to dog owners than non-owners. In a fast periodic auditory stimulation-based EEG study, we presented dogs (N=16) and humans (N=21) with continuous, rhythmically patterning vocalization streams (dog/human): every third sound differed in vocalizer identity. We compared neural entrainment to changes in human vs. dog voice identities and predicted that if experiential relevance outweighs ancient evolutionary relevance (conspecificity), then both species' brains will entrain more on human voice identity changes. We found that both dog and human brains entrained more to the rhythmically changing human vocalizer identities suggesting that they discriminated human voices more efficiently, despite that dog and human sounds were acoustically similarly discriminable. Humans overall showed entrainment to dog vocalizer change, dog owners, however, did not outperform non-owners. Experiential relevance may thus outweigh ancient preferences in voice identity processing, but individual experience with a vocalizer species does not necessarily improve identity discrimination abilities.

Introduction

Conspecific voice (identity) processing among species

Voices carry large amounts of highly important information in the everyday life of many animals. Neural pathways and brain regions supporting the processing of vocalizations have been described across species (for review, see Andics & Faragó 2018; Perrodin & Petkov 2018; Ruthig & Schönwiesner, 2022; frogs: Kelley, 2004; guinea pigs: Šuta et al., 2003, 2007, 2013; Syka et al., 2005; chimpanzees: Taglialatela et al., 2009). Several studies show that certain brain regions process voices, particularly conspecific sounds preferentially over non-vocal or heterospecific sounds (birds: Hauber et al., 2013, Pinaud & Terleph, 2008; dogs: Andics et al., 2014; rodents: Portfors et al., 2009; primates: for review, see Belin et al., 2018; marmosets: Jia et al., 2023; rhesus

macaques: Petkov et al., 2008; humans: Andics et al., 2014, Belin et al., 2000), suggesting that the vertebrate auditory system may be tuned to specific, highly relevant classes of vocal stimuli. The presence of these specializations probably contributes to the efficient processing of the diverse information voices provide. However, the primary factor (conspecificity or other) driving the emergence of these processing preferences is not yet known.

Among the plenty of socially relevant information conveyed by voices, they can provide insight into the emotional and motivational state of the animals, as well as clues to the identity of the vocalizer. The ability to discriminate voice identities is essential in many social interactions. Revealing the identity of the vocalizer can assist in identifying the caller's relationship to the receiver and aid the animal in best adjusting its behavior to the current circumstances.

Voice identity discrimination requires the animals to form categories of voices by ignoring within-category and paying attention to between-category differences. These categories can be defined both on the class level and on the individual level. Class-level voice identity discrimination occurs in many species: discriminating kins vs. non-kins (e.g., elephants: McComb et al., 2003; monkeys: Rendall et al., 1996), familiar individuals vs. unfamiliar ones (e.g., wolves: Palacios et al., 2015; dingoes: Deaúx et al., 2016; bats: Arnold & Wilkinson, 2011; ravens: Szimpl et al., 2015) females vs. males (e.g., birds: Hahn et al., 2015; humans: Weston et al., 2015), distinguishing the voices of group members vs. strangers/invaders (e.g., frogs: Bee & Gerhardt, 2002; Bourne et al., 2001; birds: Weeden & Falls, 1959).

Individual-level categorization of voices, although less explored, can also be important. Females of several species can distinguish the mating partner from other males (e.g., birds: Blumenrath et al., 2007; Lind et al., 1996; gulls: Wooller, 1978) or recognize the offspring's voice (e.g., birds: Aubin et al., 2000; seals: Petrinovich, 1974; primates: for review, see Belin, 2006). However, in these cases, there is only a single individual discriminated from all others. In species having complex social/territorial interactions or living in dominance hierarchies it can also be essential to tell several individuals apart. There are a handsome of studies showing that some birds (Kondo et al., 2010), dolphins (Janik & Sayigh, 2013) and certain primate species (Kojima et al., 2003; Petkov et al., 2008; Rendall et al., 1996) are able to individually discriminate several of their social mates' vocalizations. In addition, some species have also been reported to discriminate even unfamiliar conspecific individuals by their voice (birds: McDonald, 2012, Weary & Krebs, 1992; dogs: Molnár et al., 2009; monkeys: Briseño-Jaramillo et al., 2015, Ceugniet & Izumi, 2004).

The effect of conspecificity and experiential relevance on voice (identity) processing

The above listed studies all showed examples for conspecific voice identity discrimination (either on class- or individual-level) and, as mentioned earlier, previous works also suggest that the voices of conspecifics, typically the most relevant social partners, are processed preferentially by the brains of many species (birds: Hauber et al., 2013, Pinaud & Terleph, 2008; dogs: Andics et al., 2014; rodents: Portfors et al., 2009; primates: for review, see Belin et al., 2018; marmosets: Jia et al., 2023; rhesus macaques: Petkov et al., 2008; humans: Andics et al., 2014, Belin et al., 2000). These findings both seem to underline the importance and the special status of same-species voices in the lives of several animals. Therefore, it is possible that there is an ancient neural tuning to conspecific voices that leads to the presence of the observed (conspecific) voice-selective brain regions and the fine perception of social mates' vocalizations. However, it is unclear whether this processing preference, this sensitivity, is indeed primarily driven by conspecificity or whether experiential relevance is the key.

Intense exposure and experience with certain socially relevant vocalizations may also lead to the emergence of the observed processing preferences. There is robust literature indicating that expertise can produce both functional and structural changes in the human brain: increased gray-matter volume, extended/alternated activation or stronger/more robust functional connectivity in task-related regions were observed all over the brain in e.g., musicians (Mongelli et al., 2017), dancers (Amoruso et al., 2017), taxi drivers (Maguire et al., 2000), perfumers (Delon-Martin et al., 2013), architects (Kirk et al., 2009) compared to naïve subjects. In addition to expertise, familiarity and experience more generally also shape voice identity perception in humans: studies show that familiar voice identities are sorted more accurately than unfamiliar or lab-trained voices (Kanber et al., 2022; Lavan et al., 2021; Stevenage et al., 2020), and humans also discriminate speaker identities more efficiently when the voices are speaking in a familiar (native) language (Johnson et al., 2011). These findings all suggest that life experience may have an effect on the emergence of processing preferences in the brain.

However, adjudicating between the conspecificity and experiential relevance hypotheses is challenging. In many cases the same outcome, namely the presence of conspecific voice-selective brain regions and preferential, fine processing of certain voices, could be observed if the brain's processing preferences were for experientially relevant or anciently relevant (conspecific) sounds. This is because in many species, socially relevant partners are also conspecifics. There are only a

handsome of studies reporting heterospecific voice identity discrimination in animals living in close contact with humans, like carrion crows, horses, rhesus macaques or captive cheetahs (Wascher et al., 2012; for review, see Carlson et al., 2020; Kriengwatana et al., 2014). However, these experiments mostly demonstrate evidence for class-level (familiar vs. unfamiliar) discrimination of human voices. Also, it is not known how these animals process human voices relative to those of conspecifics, and for the mentioned species, humans are probably not the most relevant social partners.

The special interspecific relationship of dogs and humans provides an ideal study case to test the contributions of ancient evolutionary relevance (conspecificity) and experiential relevance to the shaping of processing preferences in voice identity perception. Dogs have not only been living in the social-vocal environment of humans for thousands of years (Thalmann et al., 2013), but they have also been extensively selected to communicate efficiently with humans (Miklósi, 2015), and most of them are also raised by and living with humans. Therefore, human vocalizations may carry greater social relevance than conspecific vocalizations for these dogs. The rapid evolutionary (domestication), selective and developmental history of dogs all indicate that they and their brains might have become tuned to humans, resulting in the efficient processing of heterospecific (i.e., human-originated) communicative cues, including voice identities.

Although dogs and humans may not have been equally interdependent during the domestication process and coexistence history, rapid evolutionary adaptations to changing social demands may have also occurred on the human side and shaped the sensitivity of their brain to dog vocalizations. Also, developmental, short-term experiential effects can have an impact on the processing preferences of the human brain. Particularly, dog ownership or the length of time an individual has lived with dogs may have an effect: for humans with dogs, dogs can be more relevant than for those without, and this can be manifested in how they process dog voices and, specifically, voice identities (Molnár et al., 2006). In the present study, we tested the ancient evolutionary relevance and the experiential relevance hypotheses through voice identity discrimination preferences in the dog-human relationship.

Voice identity processing in dogs, humans and within the dog-human relationship

There is a growing number of studies showing that dogs can efficiently discriminate both conspecific and human individuals by their voice. In a habituation-dishabituation experiment, Molnár and colleagues have shown that dogs can discriminate between two unfamiliar, same-breed

conspecific barks (Molnár et al., 2009). In addition, the otherwise less investigated and less understood heterospecific voice identity discrimination is becoming increasingly well described in dogs towards humans. Dogs can discriminate between male and female voices: they looked longer at the person of the correct gender when a prerecorded male or female voice was played (Ratcliffe et al., 2014). Other results suggest that dogs match the faces of their owners with their voice: they stared longer at the mismatched photos (i.e., showing a person different from the vocalizer) than when the voice and photo matched (Adachi et al., 2007). Dogs can also discriminate their owner's voice from that of an unfamiliar person as tested in behavioral studies where dogs were presented with neutral sentences either pre-recorded and displayed through loudspeakers placed behind a wall (also hiding a person) (Gábor et al., 2022) or spoken live by a hidden person (Gábor et al., 2019). In both cases, dogs chose their owner's voice significantly more often than that of control persons' (Gábor et al., 2019, 2022). It also seems that, similarly to humans, dogs use fundamental frequency and jitter to identify speakers, but they might not make use of other acoustic parameters in speech that are perceptually important identity markers for humans (e.g., harmonics-to-noise ratio, formant dispersion) (Baumann & Belin, 2010; Latinus et al., 2013; Latinus & Belin, 2011) (Gábor et al., 2022). Moreover, dogs can not only discriminate familiar voice identities or perform class-level discrimination (owner vs. non-owner), but a few studies suggest that they spontaneously discriminate between the voices of human strangers (Boros et al., 2020; Root-Gutteridge et al., 2019). Furthermore, human voice identity discrimination has also been investigated in dogs using neuroimaging methods (Boros et al., 2020; Gábor et al., 2021), but neural evidence for conspecific voice identity discrimination is completely lacking in dogs.

Humans excel at discriminating the voices of their social partners and strangers. There is evidence that infants (Mehler et al., 1978) and even fetuses (Kisilevsky et al., 2003) can discriminate the voice of their mother from that of a female stranger. At the age of 7 months, infants already distinguish unfamiliar individuals by their voice (Johnson et al., 2011). The neural basis of voice identity discrimination in humans is also well understood. Several studies have described brain regions that process changes in voice identity (Andics et al., 2013; Belin & Zatorre, 2003; Kriegstein & Giraud, 2004; for review, see Belin, 2006; Belin et al., 2018). It is also known that voices and voice identities are represented in the human brain in a prototype-centered way: listeners extract averages when forming an abstracted space of voices and use these norms to efficiently determine individual identities (norm-based coding) (Andics et al., 2010, 2013; Latinus

et al., 2013; Lavan et al., 2019). On the neuronal level, a tuning (lower spiking response) around norms, a selective suppression of average voice structures may underlie this process as described similarly for faces (Koyano et al., 2021). This coding mechanism may enable the efficient extraction of distinctive and suppression of redundant characteristics of voice identities.

It is less investigated how humans process dog vocalizations. The findings of Pongrácz and colleagues indicate that, behaviorally, humans are able to categorize barks based on the situation they were recorded in (Pongrácz et al., 2005). However, there is no experimental evidence to suggest that humans can discriminate dog identities based on their voice. In fact, Molnár and colleagues found the opposite; humans were exposed to dog bark pairs either recorded from the same or different individuals and asked to decide whether they heard one or two dogs. Humans in general could not reliably distinguish dog vocalizer identities but some of their results suggest that dog owners may perform better in the task than non-owners (Molnár et al., 2006). Neural experiments on dog voice identity discrimination in humans have never been conducted.

To sum up all the above, the reviewed studies gave examples for the efficient processing of both conspecific and relevant heterospecific voices (i.e., human voice identity discrimination by dogs). However, these were all separate experiments with varying designs and lacked the contrast of dog vs. human voice identity discrimination abilities of the two species. A direct comparison would be needed to assess possible conspecific or experiential advantages in voice identity perception and disentangle the two hypotheses. Here, to test the contributions of ancient evolutionary relevance (conspecificity) and experiential relevance to the emergence of processing preferences in the brain, we investigated the ability of dogs and humans to discriminate conspecific and heterospecific (dog/human) voice identities in a single study. To overcome some of the challenges observed in voice identity discrimination and dog brain research, we used a novel method, fast periodic auditory stimulation combined with electroencephalography, to assess spontaneous categorization of dog and human voice identities in a comparative experiment.

Fast periodic auditory stimulation for investigating voice identity discrimination

Testing spontaneous voice identity discrimination or any kind of categorization in non-human animals is not easy, as we cannot explicitly ask the animals whether they perceived something as same or different. One approach in classical behavioral experiments investigating voice identity discrimination is to use playbacks and observe the natural behavior of the animals, e.g., a mother will approach her offspring or respond with pup attraction calls when hearing the distress calls

(Petrinovich, 1974); contact calls of familiar but not unfamiliar individuals will attract mates at roosting sites (Arnold & Wilkinson, 2011). Although this method is advantageous as it measures the animals' natural behavior, it does not work well when we cannot predict any specific behavior in response to a call or when responses to different animals' calls would probably not cause great differences in the receiver's behavior (for review, see Carlson et al., 2020). To overcome this obstacle, a possible solution is to teach the animal to go for, point to, look at, etc. the correct stimuli, but this approach may not test spontaneously occurring discriminations and requires training (Ceugniet & Izumi, 2004; Weary & Krebs, 1992). Habituation-dishabituation paradigms, which measure the change in subjects' response to repeated presentation of a particular stimulus type followed by a novel stimulus, can be useful for testing spontaneous categorizations. However, it is challenging to design habituation-dishabituation test protocols that are comparable and meaningful for both human adults and non-human animals.

Fast periodic stimulation (FPS) combined with EEG provides an opportunity to capture the brain's spontaneous categorization of external stimuli objectively, directly and in a comparative way. FPS is a frequency-tagging approach that takes advantage of the phenomenon of neural entrainment, in particular entrainment by environmental rhythms. This means that when the sensory system is exposed to a (perceptually important) periodic external stimulus, this stimulation will interfere with the ongoing endogenous neural oscillations, and this endogenous brain rhythm will align and synchronize with the external rhythm (Bánki et al., 2022; Lakatos et al., 2019). In practice, we can observe that when rhythmic stimulation is combined with EEG recording, the most dominant frequency of the electrophysiological response becomes the same as that of the external stimulation (Barbero et al., 2021; Norcia et al., 2015; Vos et al., 2023). This neural entrainment can facilitate the effective processing of sensory information and be a core component underlying attentional processes (Lakatos et al., 2019; Obleser & Kayser, 2019). In an oddball FPS paradigm, there are usually a few repetitions of short stimuli from the same category (fast, base frequency) followed by a different category appearing every x^{th} presentation (slower, target frequency) (e.g., see Barbero et al., 2021; Vos et al., 2023). The key assumption is that the brain will only entrain to the slower frequency if it can discriminate between the presented stimulus types and form categories from them (Norcia et al., 2015). As a result, it provides an objective and sensitive way to investigate the categorization of the given stimuli (like voice identities), since the response is measured at a predefined frequency. Another advantage is that the method does not

require any explicit behavioral response from the subject. Therefore, it can be used in clinical and animal experiments also in a comparative manner for both human and non-human animal subjects. The independent recording from behavioral performance also helps to exclude any irrelevant attentional and decisional processes. Furthermore, FPS allows for the presentation of a large number of stimuli in a short amount of time, making it possible to investigate a given category in a representative manner within a brief period (Barbero et al., 2021), which is particularly important when conducting EEG with freely moving, awake animals. FPS has long been used with visual stimuli (for review, see Norcia et al., 2015) due to its robustness and high sensitivity. Recently, it has also been adapted to the auditory modality in humans (fast periodic auditory stimulation-FPAS) (Barbero et al., 2021; Vos et al., 2023).

Here, we presented family dogs and humans with continuous streams of short vocalizations (dog/human) in a rhythmic oddball design: two sounds from the same vocalizer were always followed by a third sound from a different vocalizer resulting in a base frequency (defined by all sounds) and a target frequency (defined by oddball sounds). We assumed that the brain would entrain to the target frequency only if voice identity discrimination occurred. We predicted that if evolutionary relevance outweighs experiential relevance during mammalian voice perception, then the brains of both species would entrain more to the conspecific stream, but if experience is the key, then neural entrainment in both species is expected to be greater to the human voice stream, and dog owners may discriminate dog identities more efficiently than humans without dogs.

Materials and methods

Participants

28 dogs and 23 human participants were measured in the experiment, 7 dogs and 1 human had to be excluded because of severe movements during the EEG recording leading to poor signal quality. The analyses were done on 21 dogs (11 female, age range 2-12, mean $\pm$ SD: 7.7 ± 3.2 , various breeds, Table S1) and 22 humans (10 female, age range 26-61, mean $\pm$ SD: 44.1 ± 11.3 , Table S2). The human participants were balanced in terms of their experience with dogs: 7 people had never had a dog, 8 currently had a dog, 7 did not have a dog at the time of the experiment, but had had in the past. All participants reported normal hearing and no neurological disorders. Dog owners and human participants volunteered to take part in the experiment in return for modest

monetary compensation and gave written informed consent. Experimental procedures met the national and European guidelines for animal care and were approved by the University Institutional Animal Care and Use Committee (UIACUC, Eötvös Loránd University, Hungary) under the certificate no. ELTE-AWC-010/2021. All procedures involving humans were approved by the Medical Research Council Scientific and Research Ethics Committee (ETT-TUKEB, Budapest, Hungary) under the certificate no. IV/2739-1/2021/EKÜ and were in accordance with the 1964 Helsinki declaration and its later amendments.

Stimuli

We presented dog barks and human speech from different individuals for the subjects. Bark samples from 4 individuals (2 male, age range 2-10 years, mean $\pm$ SD: 6 ± 3.4 , Mudi breed) were collected from a database and were recorded in a situation when the dog was left alone in a park, tied to a tree (Pongrácz et al., 2005), therefore the aim of their vocalization was probably to call the attention of their owner. The identity of these dog bark samples was known to be reliably discriminated by both dogs and classifiers (Molnár et al., 2008, 2009). As dog vocalizations do not show sexual dimorphism (Taylor et al., 2014), but human ones do, we only used the voice of one sex in the case of humans. Female individuals were selected to better match the generally high F0 mean of dog barks (range: 530-894 Hz, mean $\pm$ SD: 675 ± 107 Hz). Human voices were recorded from 4 female speakers (age range 22-30 years, mean $\pm$ SD: 26.1 ± 3.8). To mimic as closely as possible the rhythmic pattern of the dog barks, human vocalizers had to quickly utter the syllable “há” (/ha:/) several times. To also match the context of the sound recording, a dog (as a heterospecific partner) was always present during the sound recording and the vocalizers were asked to call their attention using only the given syllable.

We selected 30-30 sounds (Fig. 1) from each vocalizer of the two species and performed a permuted discriminant function analysis (DFA), which showed that both the dog and human voices were similarly highly discriminable (mean accuracy for dogs: 0.998, for humans: 0.992; all $ps < 0.001$) by mean fundamental frequency (F0), harmonics-to-noise ratio (HNR) and jitter (all known to be identity-diagnostic cues (Baumann & Belin, 2010; Gábor et al., 2022; Latinus et al., 2013; Latinus & Belin, 2011)), although the weight of the three parameters in the discrimination was slightly different for the two species. The human sample was most discriminable by F0 mean (mean accuracies: F0 mean: 0.894, HNR: 0.557, jitter: 0.420) and dog vocalizations by HNR (mean accuracies: HNR: 0.787, F0 mean: 0.674, jitter: 0.576). All sounds were < 250 ms (mean $\pm$

SD: 207.5 ± 33.9 ms), equalized in overall energy (RMS) and faded-in/-out with 10 ms ramps using Praat software (Boersma & Weenink, 2023) and Matlab R2023a (The Mathworks, Inc., Natick, Massachusetts).

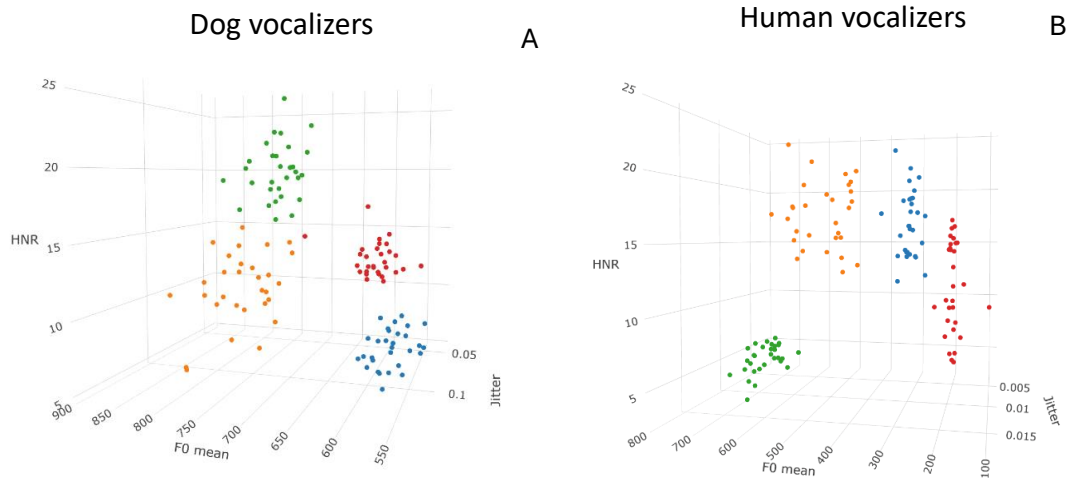


Figure 1. The acoustic distribution of the selected vocalizations.

Distribution of the 30 dog sounds (A) and the 30 human sounds (B) are depicted along HNR, F0 mean and jitter. Colors represent individuals, dots represent sound samples.

To maximize the rhythmicity of the sound presentation, we marked the so-called attack point on the waveform of each vocalization. This is the onset of the sudden amplitude increase (see Fig. 2), which marks the point at which the sound presumably becomes to be perceptually clearly salient. Each vocalization was then placed in a 250 ms long window between silent periods so that the specified point was always set at 25 ms from the start of the window.

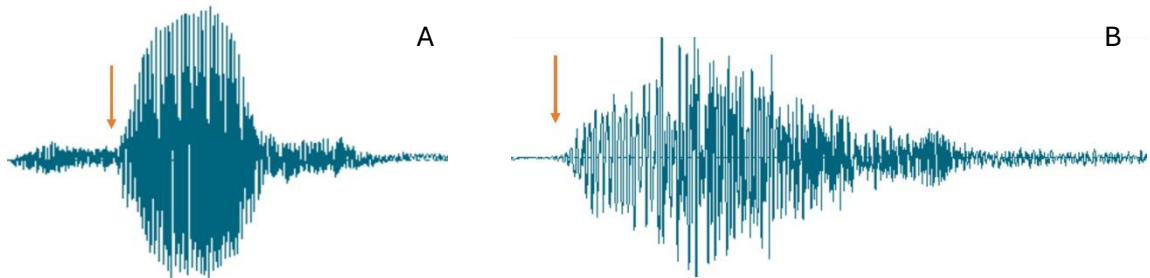


Figure 2. Attack points.

Example for a labeled attack point (orange arrow) on the waveform of a “há” syllable (A) and a bark (B).

Experimental design and stimulus presentation

During the experiment, the subjects were presented with four conditions: human and dog regular and human and dog random. Regular conditions followed a rhythmic oddball pattern: we played two sounds from the same individual (base vocalizer), while every third sound was from a different vocalizer (oddball vocalizer) in a pseudo-randomized order, with consecutive oddball vocalizers never matching in identity (e.g., AABAACAAD...). This way, a new sound was presented in every 250 ms and an oddball appeared at every 750 ms ($1/0.25\text{ s} = 4\text{ Hz}$ base, $1/0.75\text{ s} = 1.33\text{ Hz}$ target frequency) (Fig. 3).

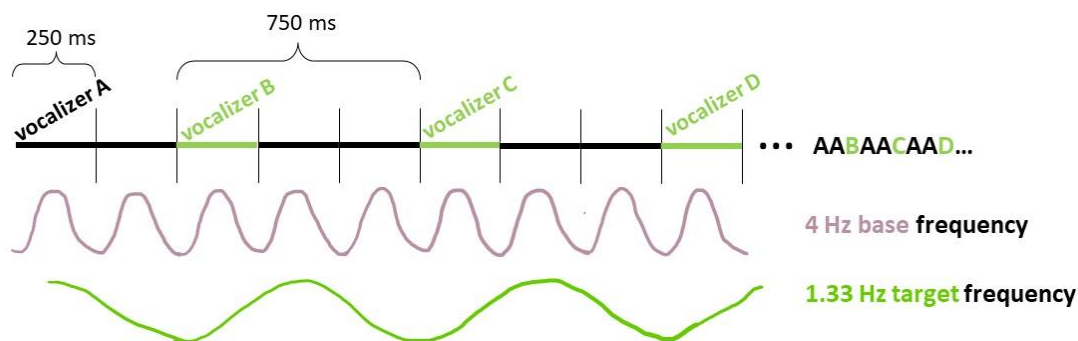


Figure 3. The rhythmic oddball pattern of sounds in the regular conditions.

Two 250 ms long sounds from the same vocalizer were always followed by a third sound from a different vocalizer yielding in a 4 Hz base frequency ($1/0.25\text{ s} = 4\text{ Hz}$) and a 1.33 Hz target frequency ($1/0.75\text{ s} = 1.33\text{ Hz}$).

The vocalizations were shown in 56.25 s long sequences, in which all the 30 sounds of the base individual were displayed, and we selected 5-5 sounds from the oddball vocalizers, each sound was repeated 5 times during the sequence to avoid entrainment caused by differences in the frequency of occurrence (De Rosa et al., 2022). The random conditions were created using the same sounds as in the regular sequences, however in this case the base and oddball sounds did not follow each other in the original AABAACAAD... pattern but in a randomized order (e.g., ABAAACADAA...) to control for possible innate/pre-existing entrainments at the target frequency that were not elicited by the presented stimuli.

Sequences of each condition were repeated 4 times during an EEG session, using every voice as base in exactly one regular and one random sequence. The order of the total 16 sequences was pseudo-randomized; presented in 4 blocks in which all the four conditions were displayed once. To reduce the risk of intersequence entrainment, we also made sure that identical conditions cannot follow each other, and added a 6.16-second sound of water lapping between sequences to interrupt entrainment (Fig. 4).

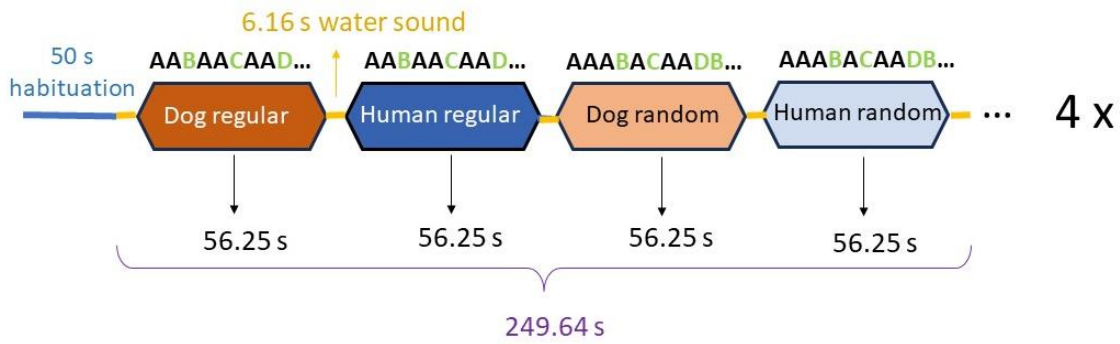


Figure 4. Schematic figure of the experimental design, the presentation of the four conditions.

The experiment began with a 50 s habituation (randomized human and dog voices from individuals other than the test identities) with 2 ms fade-in/-out to familiarize the dogs with the setup, the fast design and the barks. The habituation could be repeated several times until the dog calmed down and was also presented at least once to all human participants.

EEG recordings took place in the laboratory of the Ethology Department of the Eötvös Loránd University, which is equipped with a built-in Faraday shield in order to minimize electromagnetic noise. The dogs were always allowed to explore the room first, then they settled down with their owner either on a couch or in front of it on the floor, depending on the size and the preference of the dog. After applying the electrodes, the experimenter left the room, and the habituation sequence was repeated until the dog reached a calm lying position with its head down. The owner was always present in the room throughout the EEG recording, wearing sound-attenuating headphones. The owners were asked to put the dogs back in lying position if they stood/sat up, but not to speak and minimize any other contact with the dogs.

Human participants were seated in an armchair with their heads resting on a pillow. To minimize motion artifacts, they were asked to close their eyes, and not to move until the session finished and the experimenter entered the room.

The stimuli were played at 65 ± 10 dB from two Genius loudspeakers positioned at 170-230 cm distance in the right and left front of the subjects for both species. The auditory stimuli and the triggers to the EEG amplifier were presented by Psychtoolbox96 in MATLAB 2023a (The Mathworks, Inc., Natick, Massachusetts). The EEG sessions were video recorded for both species.

EEG acquisition

The EEG was recorded by a 68-channels Neuvo amplifier (Compumedics Neuroscan, Australia) digitized at 1 kHz sampling-rate. In case of the dog participants 8 gold-coated Ag/AgCl, surface-attached scalp electrodes fixed with Sac2 Electrode Cream were placed on the head. The placement is a modification of a canine EEG setup developed and validated in our department for sleep studies (Kis et al., 2014). The three active, cognitive electrodes were positioned over the anteroposterior midline of the skull above the bone to decrease the chance for artifacts from muscle movements (Fz, FCz, Cz). The reference electrode was placed posteriorly to the active electrodes above the inion of the occipital bone (*os occipitale*) (see Magyari et al., 2020) which is the highest point of the external occipital protuberance where the skull is usually the thickest and under which either no brain or only the cerebellum is located depending on the shape of the skull (Czeibert et al., 2019; Klaumann et al., 2018). The ground electrode was put on the left *musculus temporalis*. We also applied two eye electrodes (F7, F8) and one ear electrode (A1) to filter out muscle movements. The eye electrodes were placed on the zygomatic arch (*os zygomaticum*), in the corner of the left (F7) and right (F8) eye. The ear electrode was placed on the right *musculus temporalis* in the corner of the right ear (Fig. 5).

For human participants 64 Ag/AgCl electrodes were applied at the standard 10–20 system locations. Impedances for the EEG electrodes were kept at least below 10 k Ω , with the majority under 5 k Ω for both species.

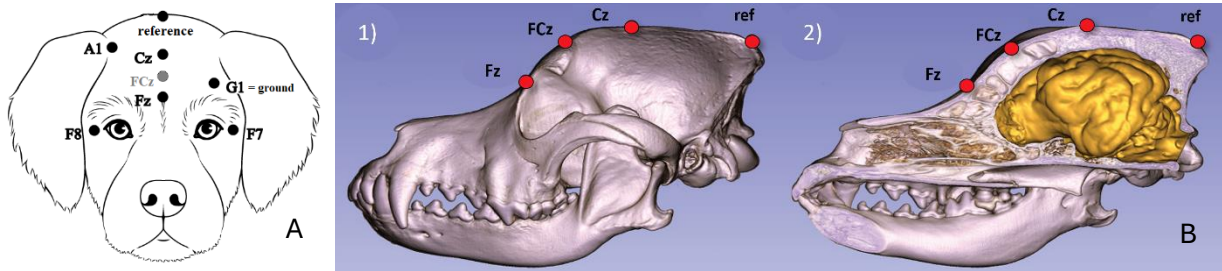


Figure 5. (A) Placement of the eight electrodes on the dog head. (B) Relative position of the 3 cognitive (Fz, FCz, Cz) and the reference electrode to the skull and the brain of a border collie (as an example of a dog with mesocephalic head type): (1) Volume-rendered CT scan of the skull, left lateral view. (2) Volume-rendered CT scan of the skull, cut in the midsagittal plane, together with the brain. (Adapted from: Magyari et al., 2020)

Data analysis

Data analysis was performed in Matlab R2023a (The Mathworks, Inc., Natick, Massachusetts), using the FieldTrip toolbox (Oostenveld et al., 2011; Donders Institute for Brain, Cognition and Behaviour, Radboud University, the Netherlands. See <http://fieldtriptoolbox.org>), R statistical environment (<https://www.r-project.org/>, v. 4.2.3, R Core Team, 2023) and RStudio software (<https://rstudio.com/>).

Preprocessing

The data were digitally filtered offline with 0.1 Hz high-pass and 40 Hz low-pass filter and segmented into the 56.25 s sequences. We cleaned the data based on visual inspection and excluded the sequences that were highly contaminated with artifacts (Barbero et al., 2021) keeping only those subjects for further analysis who had at least 7 “clean” sequences out of the 16 and at minimum one presentation of each condition survived the cleaning procedure (21/28 dogs with average 12.19 kept sequences/subject, 22/23 humans with 16 kept sequences/subject).

Frequency domain analysis

For the frequency domain analysis, the sequences from each condition were averaged separately in the time domain for each participant to attenuate EEG activity not elicited by the auditory stimulation but by internal or external noise sources or being contaminated by minor motion artifacts. A fast Fourier transformation was then performed to isolate responses at the required frequencies for each condition and each channel; the power spectra ranged from 0 to 9 Hz and had a 0.0178 Hz resolution (1/56.25 s), which allowed the fine isolation of responses at the frequencies of interest and at their harmonics. To determine the significant frequency bins, we computed the grand average power across subjects for each condition separately. Then, we calculated signal-to-noise ratio (SNR, relative power) at each frequency bin (x) by dividing the given power value by the baseline mean: average power of the surrounding 20 frequency bins (excluding the directly adjacent and the local minimum and maximum values) (Barbero et al., 2021; Retter & Rossion, 2016). As a last step, a corresponding Z-score value was computed ($Z\text{-score} = (x\text{-baseline mean}) / \text{baseline standard deviation}$) to define whether power at frequencies of interest was significantly higher than at the surrounding frequencies (thresholded at: relative Z-score > 2.3, i.e., $p < 0.011$, one-tailed, signal > noise). SNR and Z-score values, similarly to the power spectra, were calculated for each channel separately. Group-level scalp topography maps

were generated for humans using Z-score values of the target frequency at each channel to show the allocation of the responses across the head for each condition.

To further clean our data, and assess a degree of attention in the subjects, we calculated a general 4 Hz entrainment to the stimuli in all participants. We pulled all conditions together, averaged all the cleaned sequences/participant in the time domain and then calculated power spectrum, SNR and Z-score values for each participant for this averaged data. We only included those participants in any of the group-level analyses who showed a significant entrainment to the base frequency (increased power at 4 Hz compared to the surrounding frequency bins) in this averaged condition. The reported results are all based on the data of these subjects (16 dogs (9 female, age range: 2-11, mean $\pm$ SD: 6.9 ± 3.0); 21 humans (10 female, age range: 26-61, mean $\pm$ SD: 44.4 ± 11.5)).

We predicted entrainment (increased power at the frequency of interest compared to the surrounding 10-10 frequency bins) to the base frequency (4 Hz) in both the regular and random conditions in both species' brain. We expected entrainment to the target frequency in both the dog regular and human regular conditions in both species. No 1.33 Hz entrainment was expected in random conditions.

Statistical analysis

To quantify the responses and compare the magnitude of the entrainment to the target frequency in the different conditions, a baseline subtracted (absolute) power spectra was also computed separately for each participant and each channel. This baseline correction prior to the quantification and comparison aids to account for the fact that, due to the nature of the power calculation, the power will always be higher at lower frequencies and that different frequency bands are differentially affected by noise in EEG recordings (Luck 2014). We compared entrainment to the target frequency (1.33 Hz) across the conditions using linear mixed-effects models in R (LMMs, R package 'lme4', Bates et al., 2015), with condition (regular/random) and species (dog voices/human voices) as fixed effects, including their interaction and adding subject ID as a random term. Forward stepwise procedure was used to compare models based on likelihood ratio tests (LRTs). We report χ^2 , the corresponding p-values of LRTs and parameter estimates ($\beta \pm$ SE) for significant explanatory variables. Separate models were built for each channel in both species. As power increase was not always present at the harmonics in the conditions where we

found a significant 1.33 Hz effect, conditional differences were always calculated for power values at the target frequency only.

To assess the influence of ownership on humans' ability to discriminate dog vocalizations, we built further LMMs for those electrodes in humans, where we observed significant 1.33 Hz entrainment for the dog regular condition and the difference for the dog regular vs. dog random conditions was also significant. In this model, data points contained the absolute power values at 1.33 Hz elicited only by the dog regular and dog random conditions. Condition (regular/random) and ownership (currently have (now)/have had in the past (past)/never had (never)) were added to the model as fixed effects, including their interaction.

We predicted that if experiential relevance plays a greater role than evolutionary relevance in shaping voice identity processing preferences, then entrainment to human voice identity changes in the dog brain would be higher than to dog voices. If, on the contrary, evolutionary relevance outweighs experiential relevance, then dogs would show greater entrainment to the target frequency in the dog regular than in the human regular condition. In case of the human subjects, both the experiential relevance and evolutionary relevance hypothesis would be supported by greater 1.33 Hz entrainment to human than dog voice identity changes. Support for the experiential relevance hypothesis would be indicated by dog owners demonstrating greater entrainment to the dog regular voice stream than non-owners.

Results

Dogs

We observed significantly higher response at 4 Hz (base frequency) than at the surrounding frequency bins in every condition on all the cognitive electrodes (Fz, FCz, Cz) (all Z-scores > 3.92, all p s < 0.001) (Table S3).

There was a significant response at the target frequency (1.33 Hz) in the human regular condition on Fz and FCz electrodes. No significant 1.33 Hz response was found in the dog regular and the random conditions (Table 1).

Table 1. Target frequency (1.33 Hz) responses in dogs in the different conditions.

Condition	Z-score - Fz	p-value - Fz	Z-score - FCz	p-value - FCz	Z-score - Cz	p-value - Cz
Dog regular	-1.27	0.102	-0.46	0.323	0.26	0.397
Dog random	0.80	0.212	0.29	0.386	1.13	0.129
Human regular	2.30	0.011	2.33	0.010	0.64	0.261
Human random	-0.72	0.236	0.83	0.203	1.39	0.082

Condition	Z-score - F7	p-value- F7	Z-score - F8	p-value - F8	Z-score - A1	p-value - A1
Dog regular	-0.95	0.171	-1.03	0.152	0.57	0.284
Dog random	-0.16	0.436	1.05	0.147	1.02	0.154
Human regular	1.23	0.109	1.11	0.134	0.92	0.179
Human random	-1.99	0.023	-0.74	0.300	-0.30	0.382

Linear mixed-effect models comparing the magnitude of the 1.33 Hz entrainment in the different conditions revealed a significant interaction between species (dog/human) and condition (regular/random) on Fz electrode ($p = 0.043$). Post-hoc pairwise comparisons revealed a trend of the baseline-subtracted power values in the human regular condition being higher than in the human random condition (human regular vs. human random: $\beta \pm SE = 0.23 \pm 0.12$, $t = 1.96$, $p = 0.056$). A similar trend-level effect was observed between human regular and dog regular conditions (dog regular vs. human regular: $\beta \pm SE = -0.22 \pm 0.12$, $t = -1.90$, $p = 0.064$). There was no significant difference in the dog random vs. human random ($\beta \pm SE = 0.12 \pm 0.12$, $t = 1.05$, $p = 0.299$) or the dog regular vs. dog random ($\beta \pm SE = -0.11 \pm 0.12$, $t = -0.97$, $p = 0.335$) comparisons (Fig. 6). No interaction or main effect was found on the other channels (FCz, Cz, F7, F8, A1, for details, see Table S4). Figure 7 represents the SNR spectra on Fz electrode for the different conditions.

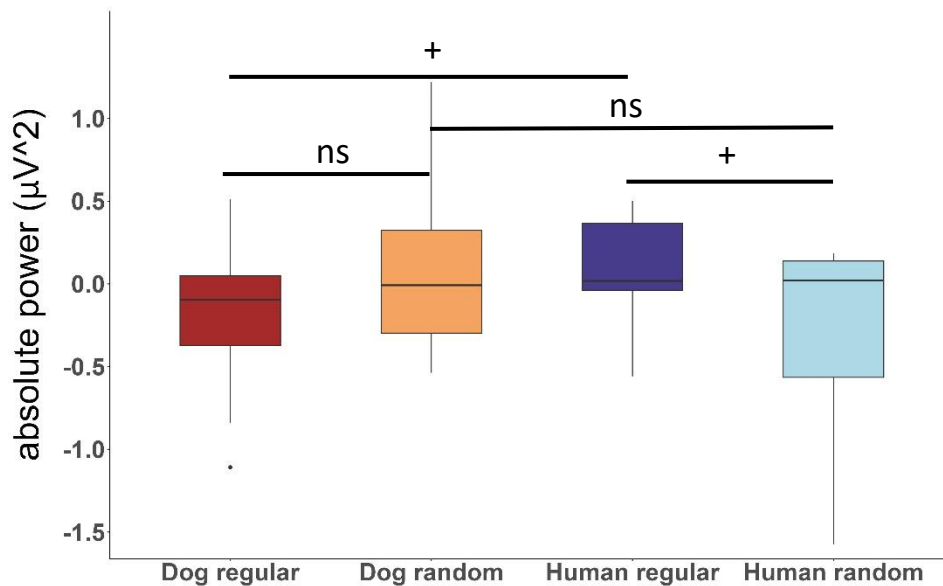


Figure 6. Target frequency responses in dogs for the different conditions.

Responses are baseline-subtracted (absolute) power values on Fz. Boxplots indicate the median, 25th and 75th percentiles (boxes), and the minimum and maximum (whiskers) values, +: trend-level difference, $p < 0.065$.

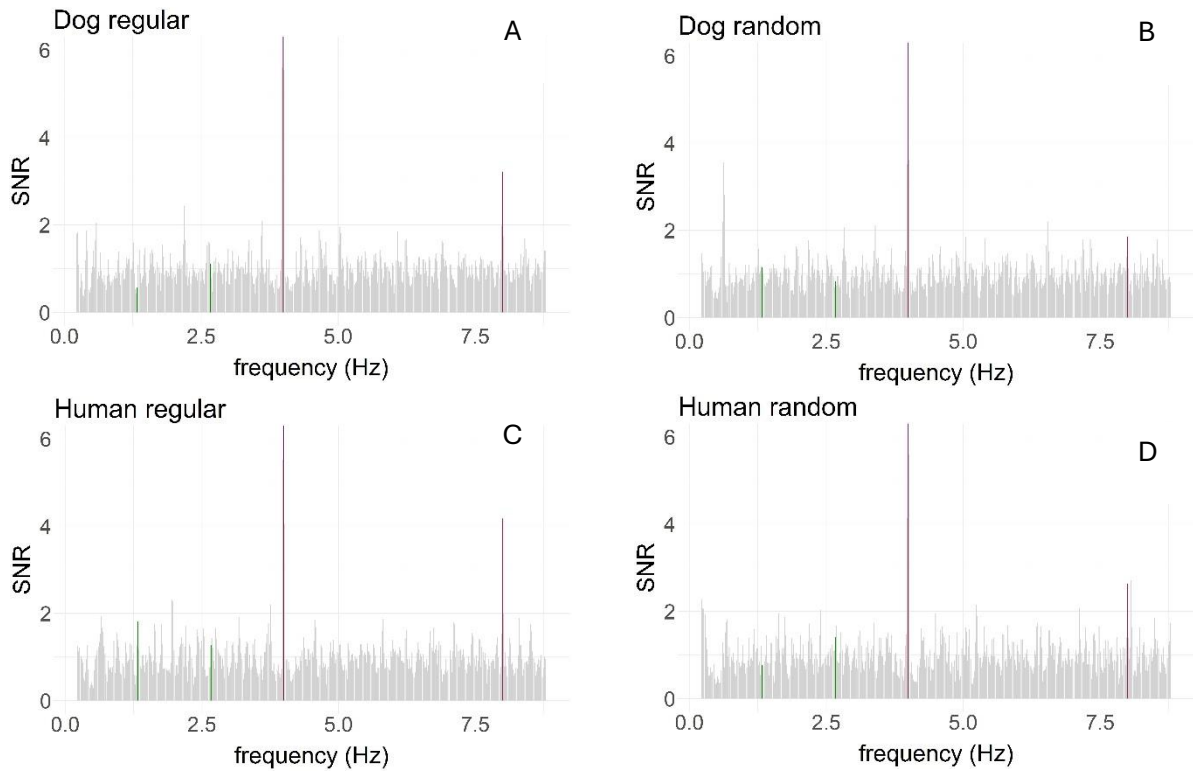


Figure 7. SNR (signal-to-noise ratio) spectra in dogs for the different conditions on Fz electrode
(green: 1.33 Hz and its first harmonic (2.66 Hz), purple: 4 Hz and its first harmonic (8 Hz)).

Humans

Significantly higher response at 4 Hz (base frequency) than at the surrounding frequency bins were found in every condition on almost all the electrodes, see Figure S1.

We observed significant responses at the target frequency (1.33 Hz) in the dog regular condition mainly on occipital, parietal, occipito-parietal electrodes and on a few frontal electrodes (all Z-scores > 2.52, all p s < 0.001). Significant target frequency responses were shown in the human regular condition on occipital, parietal, occipito-parietal, centro-parietal, temporal and frontal electrodes (all Z-scores > 2.45, all p s < 0.001). No significant 1.33 Hz response was found for the dog and human random conditions on either of the electrodes (Fig. 8).

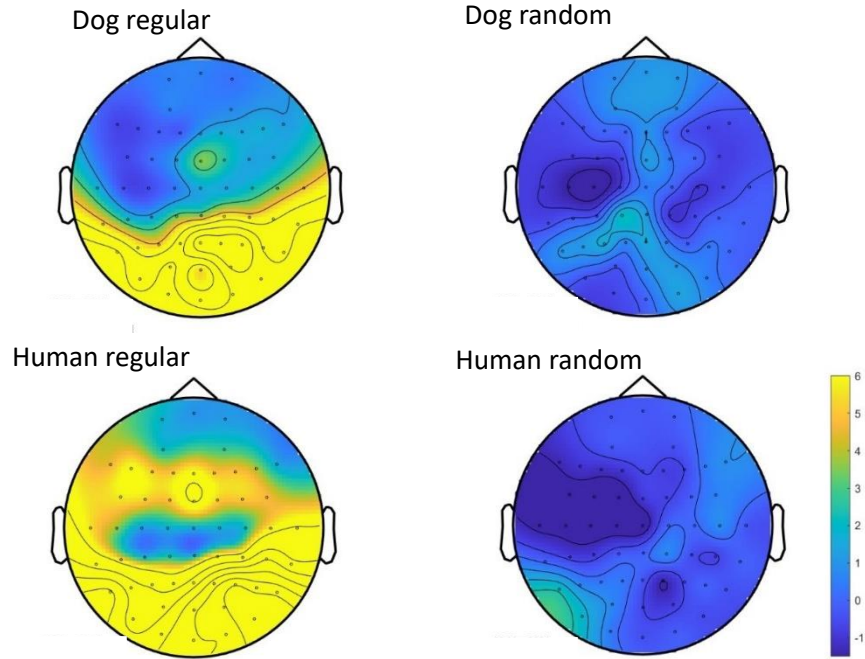


Figure 8. Topographic scalp maps for target frequency responses in humans for each condition.

The scale of the plots represents Z-score values for power increase at the target frequency (1.33 Hz) relative to the surrounding 20 frequency bins, which were considered to be significant above 2.30 ($p < 0.011$, one-tailed, signal > noise).

Linear mixed-effect models comparing the magnitude of the 1.33 Hz entrainment in the different conditions revealed a significant interaction between condition (regular/random) and species (dog voices/human voices) on 26 channels (all $ps < 0.037$) at occipital, parietal, occipito-parietal and frontal regions (for details, see Tables S5, S6). The corresponding post-hoc pairwise comparisons showed that baseline-subtracted power values at 1.33 Hz were significantly higher in the human sequences than in the dog sequences for the regular conditions on all 26 channels (dog vs. human: all $ts < -2.12$, all $ps < 0.038$, Fig. 9.A, for details, see Table S7.C), but not for the random conditions on either of the channels (dog vs. human: all $ts < 1.11$, all $ps > 0.272$, for details, see Table S7.D). Also, both dog and human regular conditions elicited significantly higher power values at 1.33 Hz than their random pairs on several channels (human regular vs. human random: all 26 channels (Fig. 9.B), all $ts > 3.44$, all $ps < 0.001$; dog regular vs. dog random: 16 occipital, parietal and occipito-parietal channels (Fig. 9.C), $ts > 2.11$, $ps < 0.039$) (for details, see Table S7.A and B) (Fig. 10). Figure 11.A and B depict the averaged SNR spectra across those channels, which showed a significant difference in the dog regular vs. dog random comparison. Figure 11.C and D

represent the averaged SNR spectra across the significant electrodes in the human regular vs. human random comparison. As shown on this Figure 11, power was significantly increased at 1.33 Hz compared to the surrounding frequency bins on the average of the beforementioned electrodes in the dog regular (Z-score = 9.22, $p < 0.001$) and human regular (Z-score = 23.77, $p < 0.001$) conditions but not in the random conditions (dog random: Z-score = 0.39, $p = 0.348$; human random: Z-score = -0.27, $p = 0.394$).

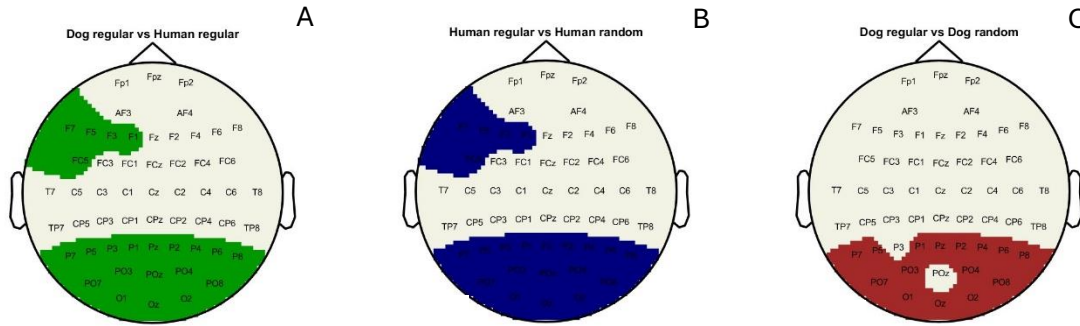


Figure 9. Topographical distribution of the channels showing a species-condition interaction.

The topoplots depicts the results of the post-hoc tests: channels showing significant difference in the baseline-subtracted power values at 1.33 Hz for the dog regular vs. human regular (A), human regular vs. human random (B) and dog regular vs. dog random conditions (C).

Out of the other channels not showing a significant interaction between species and condition, 29 central, centro-parietal, temporal, frontal and centro-frontal electrodes revealed a significant effect of condition (all p s < 0.04). At all of these channels the baseline-subtracted power values at 1.33 Hz were higher for regular than for random sequences (all t s > 2.10 , all p s < 0.04) (Fig. 12, for details, see: Tables S8, S9).

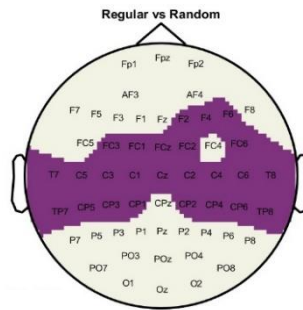


Figure 12. Topographical distribution of the channels showing a condition effect.

Colored channels showed no interaction between species and condition, but a significant condition effect: the baseline-subtracted power values at 1.33 Hz being higher for the regular than for the random sequences.

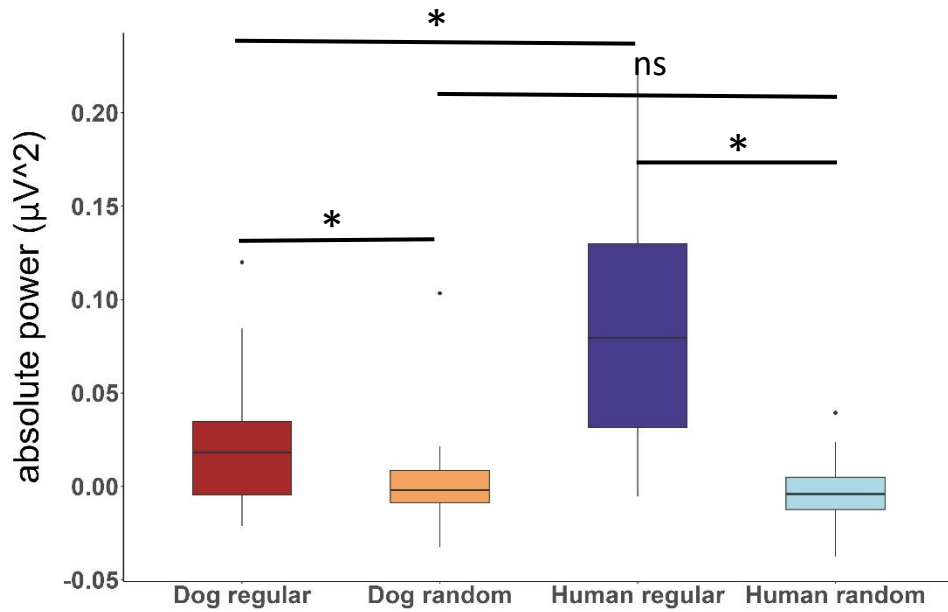


Figure 10. Target frequency responses in humans for the different conditions.

Responses are the average of baseline subtracted (absolute) power across the channels showing significant interaction between species and condition. Boxplots indicate the median, 25th and 75th percentiles (boxes), and the minimum and maximum (whiskers) values. The models were calculated for each channel separately as reported in the text, the asterisks in the figure are just for illustrative purposes, *: all ps < 0.05.

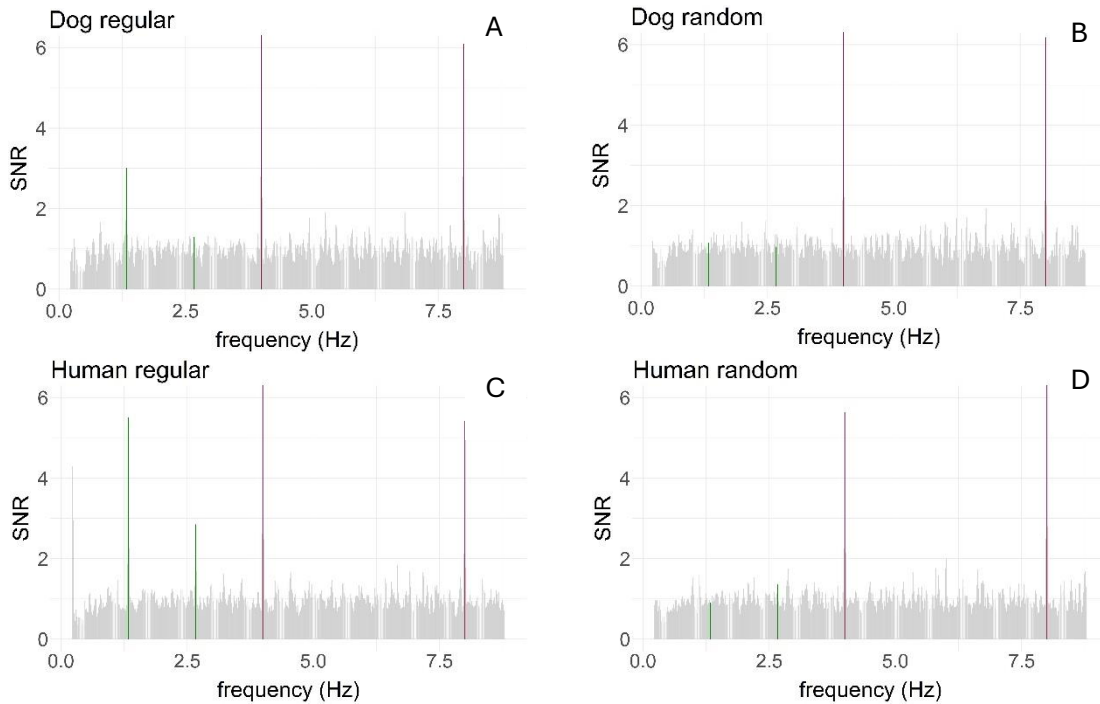


Figure 11. SNR (signal-to-noise ratio) spectra in humans for the different conditions.

Responses are averaged across the electrodes showing significant differences in the dog regular vs. dog random (A, B) or the human regular vs. human random comparisons (C, D) (green: 1.33 Hz and its first harmonic (2.66 Hz), purple: 4 Hz and its first harmonic (8 Hz)).

We found significant interaction between condition (dog regular/dog random) and ownership (now/past/never) on only a single channel (C6) out of the 29 showing a significant 1.33 Hz entrainment for the dog regular condition and a difference in absolute power values at 1.33 Hz for the dog regular vs. dog random comparison ($p = 0.009$) (for details, see Tables S9, S10). The post-hoc test for this single channel (C6) revealed significantly higher absolute power for the dog regular condition in the group never owning a dog than in the group currently having a dog ($t = 3.12$, $p = 0.015$). As expected, for the dog random condition no difference in the 1.33 Hz absolute power values was observed for any of the ownership group comparisons (all t s < 1.00 , all p s > 0.588).

Discussion

Here, in a comparative EEG experiment, we investigated dogs' and humans' ability to discriminate con- and heterospecific voice identities. We found that human brains entrained to both human and dog voice identity changes showing greater power increase at the frequency of the rhythmic voice identity change in human than dog sound streams, i.e., better discrimination for human voices. Dog brains also entrained to the target frequency in the regularly patterned human sequences but did not show evidence for the discrimination of dog vocalizer identities: no power increase at the target frequency of the dog voice identity change was observed. Together, these findings provide the first EEG evidence for voice identity discrimination in dogs and demonstrate humans' ability to discriminate dog identities based on their voice alone.

We successfully applied a new method, fast periodic auditory stimulation combined with EEG to demonstrate evidence of voice identity discrimination in dogs and humans. Increased power at the target frequency (1.33 Hz) was seen in the regular conditions meaning that the dog and human brain entrained to the voice identity changes (dogs to human voices, humans to both dog and human voices). These findings suggest that both species could form categories of voice identities (same or different individual as the base vocalizer), and discriminate vocalizers in a fast design. Furthermore, as expected, dog and human brains entrained to the target frequency only in the regular conditions, indicating that the 1.33 Hz rhythm was indeed elicited by the presented stimuli and was not present due to an endogenous brain rhythm. This experiment is one of the first to show that the dog brain, similarly to the human brain (Barbero et al., 2021; De Rosa et al., 2022; Vos et al., 2023), can entrain to rhythmic external stimuli. In sum, these findings suggest that fast periodic

stimulation combined with EEG can be a powerful method for future comparative experiments on categorical discrimination abilities, as it can be applied to both human and non-human animals with ease. All the more so, since fast periodic stimulation enables the presentation of a large number of stimuli, yielding robust data in a short amount of time, which can be a crucial point in animal research involving freely moving individuals.

The finding that both dogs and humans showed an increased power at the target frequency compared to the surrounding frequency bins for the human regular condition indicates that both species' brain entrained successfully on human voice identity changes. Thus, we provided the first electrophysiological evidence that dogs, similarly to humans, can discriminate unfamiliar human individuals based on their voice alone. Multiple studies have shown that dogs discriminate their owner's voice from that of an unfamiliar person (Adachi et al., 2007; Gábor et al., 2019, 2021, 2022), but their ability to perform a more difficult discrimination and tell the voices of strangers apart is less investigated. Root-Gutteridge and colleagues revealed in a behavioral habituation-dishabituation study that dogs spontaneously categorize the same unfamiliar speaker across different words and habituate to the voice (Root-Gutteridge et al., 2019). We corroborated this evidence for identity discrimination of unfamiliar human voices in dogs and demonstrated the possibility of replacing a less comparable and difficult to implement habituation-dishabituation setup with a fast and data-rich EEG paradigm. Boros and colleagues investigated the neural correlates and localization of unfamiliar human voice identity processing in the dog brain using fMRI. They observed a repetition enhancement effect, that is increased brain activity when unfamiliar voice identities were subsequently repeated within a block compared to when the vocalizers were always changed (Boros et al., 2020). In line with this, we provided further neural evidence for dogs' ability to discriminate unfamiliar human voice identities. However, while the observed repetition enhancement is a more indirect way of demonstrating voice identity processing, as it does not directly compare the responses to individual sounds but measures differences in the activity elicited by blocks of (identity-) repeated vs. changed stimuli, the present study provides a more explicit demonstration of dogs' ability to detect identity changes in consecutive voices. It is also important to note that both in the study of Boros and colleagues and in that of Root-Gutteridge and colleagues, male and female voices were mixed during stimulus presentation (Boros et al., 2020; Root-Gutteridge et al., 2019). Therefore, voice identity differences might have been more salient, or their findings may have partially reflected discrimination of voice

genders rather than of individual identities. Since in our experiment we only used female voices, voice gender discrimination cannot account for the present findings. Thus, we demonstrated that dogs are capable of identifying unfamiliar human voices even from a much more homogenous sample.

Increased power in human EEGs at the target frequency for the dog regular condition demonstrates for the first time that humans are capable of identifying individual dogs based on their vocalizations. No neural experiments have investigated the ability of humans to differentiate between dog vocalizers, and the single previous behavioral study found no evidence that humans can reliably decide whether consecutive dog barks came from the same or different individuals (Molnár et al., 2006). This difference in measurement modality between the previous and the present experiment may be critical to the results. In particular, some studies suggest that behavioral experiments using performance measures often fail to reveal certain cognitive abilities that can be demonstrated neurally (Boros et al., 2021, 2024). Thus, these differences may contribute to the apparently contradictory results of the previous behavioral and the current neural experiment.

Greater entrainment to the target frequency in the human than in the dog regular condition suggested more efficient discrimination of human than dog vocalizer identities in both species. This human preference in humans confirms previous findings of conspecific voices being preferentially processed in the human brain (Andics et al., 2014, Belin et al., 2000). Nonetheless, it may not help to reveal the contribution of evolutionary relevance and experiential relevance to the emergence of processing preferences in voice identity perception, as for humans the most relevant social partners are probably the conspecifics. However, the finding that the tested dogs, living in human families, also tended to better discriminate human than dog vocalizer identities may indicate a greater role of experiential relevance than ancient evolutionary relevance in voice identity perception. This processing preference for experientially relevant voices is consistent with previous findings indicating that familiarity and experience affect voice identity perception. Particularly, listening to familiar individuals or vocalizers speaking in their native language both enhanced humans' performance in voice identity discrimination (Johnson et al., 2011; Lavan et al., 2021; Stevenage et al., 2020). While in these human examples expertise reflects developmental changes, in the case of dogs, the shown preference for human sounds may root both from developmental and rapid evolutionary adaptations to changing social demands.

One may argue that the shown tendency of dogs to better discriminate human than dog vocalizations does not reflect a human preference. In particular, finding no entrainment to dog, unlike to human, voice identity change in dogs may suggest that dog barks simply do not carry enough identity information to be discriminated acoustically or perceptually. However, this assumption is unlikely for several reasons. First, the same set of dog stimuli was discriminated by humans in the present experiment. Second, former studies have proved that the same short barks can be discriminated by classifiers with high accuracy (Gómez-Armenta et al., 2024; Larrañaga et al., 2015; Molnár et al., 2008). Most importantly, the results of a previous behavioral study by Molnár and colleagues indicate that dogs also have the ability to discriminate conspecific barks (Molnár et al., 2009). The reason why they found voice identity discrimination, and we did not, may reflect some differences in the complexity of the study designs. Although both experiments used the prerecorded barks of the same Mudi dog individuals, Molnár and colleagues always played a bark sequence from a single individual and then another from a different dog, whereas we presented a sequence of short, single barks from different individuals. Thus, it is possible that in the previous behavioral experiment, dogs were able to extract more identity-diagnostic information from the stimuli than they could from a single bark. It is also worth noting that in the behavioral experiment a subject always had to discriminate the voices of only two dogs in a habituation-dishabituation paradigm. In contrast, in our design dogs had to discriminate the base vocalizer from 3 different voice identities and generalize across the base sounds to show entrainment to the target frequency. Importantly, although these differences both suggest that the current setup may have been more difficult for dogs, contributing to finding no evidence for dogs' conspecific voice identity discrimination, we cannot say that the applied FPAS paradigm is not suitable to investigate voice identity discrimination in dogs. We demonstrated that in the very same design, presenting similarly short vocalizations in the same AABAAC pattern, thus requiring the same generalization and complex discrimination, dogs were able to distinguish human voice identities. This finding further supports the idea that these dogs may have become tuned to human voices, and that this expertise may have facilitated the emergence of processing preferences in their brains, resulting in the observed efficient processing of human voice identities.

Although we showed that dog vocalizations were similarly highly discriminable to the human voices based on a DFA, we cannot be certain that the acoustic parameters driving the discrimination in our acoustic measures are the ones carrying the most identity-diagnostic information in dog

voices for dogs, which can contribute to the detected poorer dog vocalization discrimination of the dog brain. DFA showed that dog and human voices were similarly discriminable by F0 mean, HNR and jitter. However, the weights of the three parameters in the discrimination for the two species were slightly different. Human stimuli were most discriminable by F0 mean, while HNR was the driving factor in dog vocalizations. We know that dogs use F0 mean and jitter to identify human vocalizers (Gábor et al., 2022), and that all these three parameters are perceptually important identity cues for humans (Baumann & Belin, 2010; Latinus et al., 2013; Latinus & Belin, 2011). However, Gábor and colleagues found no evidence that dogs use HNR when they have to recognize human individuals (Gábor et al., 2022), therefore it is possible that HNR is perceptually not that important for the discrimination of dog identities either. In contrast, Larrañaga and colleagues showed that the best performing classifiers use F0 mean and HNR among others to identify dog barks (Larrañaga et al., 2015). HNR was also part of the optimal feature subset for dog bark identity classification in the study of Molnár and colleagues (Molnár et al., 2008). In addition, the behavioral evidence that dogs orient faster to pups with noisier separation calls (Lehoczki et al., 2020) indicates that they can sense and rely on HNR variations. Furthermore, the tonality of sounds is also known to affect humans' success rate of discriminating dog vocalizations: they perform better when they have to tell apart dog barks differing in HNR than when their task is to distinguish two high HNR voices (Molnár et al., 2006). To sum up, although we cannot exclude that the harmonic-to-noise ratio of dog vocalizations is not the acoustic parameter based on which dogs best discriminate individuals, there is clear evidence for the identity-diagnostic and perceptual importance of HNR in dog vocalizations.

Although we hypothesized that dog owners would be more efficient in discriminating dog vocalizations than non-owners, we found no effect of ownership on humans' ability to distinguish dog voice identities suggesting that life experience with dogs may not have a strong influence on humans' performance. However, this does not necessarily mean that experiential relevance cannot play a role in the emergence of processing preferences in the brain. First, even if ontogenetic effects may not further improve humans' performance in dog voice identity discrimination, it is still possible that experiential relevance has played a role on a rapid evolutionary scale. Specifically, it can be that the 18-30 thousand years of the dog-human coexistence (Thalmann et al., 2013) has shaped the human brain's processing preferences and humans have generally become good at processing dog vocalizations. In support of this, humans in the present study were not only capable

of discriminating dog identities overall, but they distinguished the voices of same-breed, thus similar-sized dogs suggesting that they could detect fine differences in dog barks. To see whether this remarkable performance was indeed due to a dog preference and not reflecting a general expertise of humans to categorize voice identities, it would be favorable to investigate how humans perform in discriminating the voice identity of a socially irrelevant heterospecific species. The possible finding that humans show better discrimination of dog vocalizer identities than that of irrelevant heterospecifics would indicate the importance of the dog-human relationship from the human perspective too. Also, it would provide another argument in favor of the experiential relevance hypothesis. In contrast, the absence of an identity discrimination preference for dog voices over socially irrelevant heterospecific vocalizations would explain our current findings of humans' ability to discriminate dog identities as a general human expertise in forming categories based on acoustic input. Future research should be done to directly test this question.

The reason behind the lack of an ownership effect may also be that ownership is not the most important aspect in humans' expertise with dogs. The findings of Molnár and colleagues are not conclusive either about the effect of ownership on dog vocalizer discrimination in humans. In one experiment they found that owners could better decide whether they heard one or two dogs barking (but none of them passed the empirical threshold of reliable discrimination), while in another test no ownership effect was found on performance. Also, humans did not show improvement in the dog voice identity discrimination task when they had to distinguish their own dog's bark from that of an unfamiliar individual than when two unfamiliar barks were played (Molnár et al., 2006). These inconclusive results, together with our findings suggest that ownership may not function as a powerful measure of humans' dog expertise.

The finding that the very same set of dog stimuli that could not be discriminated by dogs was distinguished by humans may be explained by the fact that humans use a wide range of acoustic parameters for the efficient discrimination of voice identities including F0 mean, jitter, HNR and dF (formant dispersion) (Baumann & Belin, 2010; Latinus et al., 2013; Latinus & Belin, 2011). Therefore, it is possible that humans' big repertoire of perceptually important identity-diagnostic features drives their expertise in voice identity discrimination and enabled the entrainment to both human and dog voice identity changes in the present experiment. However, the cited studies all investigated which parameters contribute to the discrimination of human voices and no experiments have been conducted to identify the perceptually important acoustic features in dog

barks for individual recognition, neither with dog nor with human listeners. Therefore, further research is needed for the evaluation of the assumption that humans may rely on more acoustic cues than dogs when processing dog voice identities.

The observed variance in the topography of the dog and human responses in human listeners suggests that differences in electrode placement may also be crucial in why we detected target frequency entrainment for the dog regular condition in humans but not in dogs. Specifically, in the human brain absolute power differences at the target frequency for the dog regular vs. random condition were found mainly in occipital, parietal, centro-parietal and occipito-parietal regions, while the human regular vs. human random contrast was also significant at frontal electrodes. Similarly, the 1.33 Hz entrainment for the human regular condition was seen on Fz and FCz electrodes for dogs and the trend-level difference between human regular and human random sequences was shown at Fz electrode only. However, although the dog brain is much smaller than the human brain and a single electrode on the dog head is likely to detect the activity from a larger brain region than a single electrode on the human head, the coverage of the dog brain with the current set of 3 cognitive electrodes is probably not the same as of the human brain with the 64 electrodes used. It is especially true for the parietal and occipital lobes as shown in Figure 5. Therefore, with better coverage of the posterior region where the entrainment to the dog identity changes was mainly seen in humans, we might have detected a similar response in the dog brain as well. We may consider placing more electrodes on the dog head in the future, nonetheless, the thick layers of the *musculus temporalis* (Evans & De Lahunta, 2012) make it challenging to measure brain activity with high signal quality in posterior regions.

The topographic distribution of the human responses in our experiment mostly corroborates previous findings investigating specific aspects of voice identity discrimination in humans. Hausfeld and colleagues found better speaker discrimination in overlapping regions: at right occipito-parietal and right temporo-parietal channels (Hausfeld et al., 2012). A study by Hu and colleagues aimed to decode speaker identities based on the listener's EEG response and showed that frontal and parietal lobes provided the best recognition accuracy (Hu et al., 2021). Studies investigating class-level voice identity discrimination have also found similar topography. ERP responses for male and female voices were significantly different in posterior and central regions (Latinus & Taylor, 2012). Mean amplitudes for intimately familiar and unfamiliar voices were also found to be different at frontal, centro-frontal, centro-parietal, occipito-parietal and parietal

electrodes for late time windows (Plante-Hébert et al., 2021). The results of Zäske and colleagues are more consistent with the electrodes that generally differentiated between regular and random sequences but did not show a species-condition interaction in our experiment. They demonstrated negative power differences for unfamiliar but previously heard vs novel voice identities at central and right temporal electrodes (Zäske et al., 2014). Overall, the regions identified in the current experiment to be involved in voice identity processing are consistent with previous findings.

As detailed above, the current experiment has some limitations, such as the lack of inclusion of socially irrelevant heterospecific voices and the poorer coverage of the dog brain than the human brain with electrodes. Although these limitations do not question the main findings of the study, further research would be needed to test whether the demonstrated ability of humans to make fine distinctions between dog bark identities reflects their dog vocalization expertise or rather their general capacity to form categories based on acoustic input. Also, it would be beneficial for future dog EEG experiments to see if some modifications (e.g., applying more electrodes) to the current placement of the electrodes could achieve better coverage of the dog brain, making the measurements more sensitive to changes in neural activity and thus allowing a more precise investigation of dogs' cognitive mechanisms.

Conclusion

Using a novel method, fast periodic auditory stimulation combined with EEG, we constituted the first bits of electrophysiological evidence that dogs can discriminate unfamiliar human voice identities, and that humans can detect individual changes in dog voices. The finding that dogs seem to be more efficient in detecting voice identity changes in human than in dog voices suggests that experiential relevance may play a crucial role in forming voice identity perception, outweighing ancient processing preferences. That dogs process human voices so effectively and that humans can detect fine changes in dog voice identities but that dog owners do not outperform non-owners in dog voice discrimination, together raises the possibility that experiential relevance shapes voice identity processing on a rapid evolutionary scale rather than during development. Together, these findings shed light on the adaptive capacity of the mammalian brain to adjust its processing mechanisms to the changing social environment.

I hereby declare that I did not use AI to produce this work.

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>
- Amoruso, L., Ibáñez, A., Fonseca, B., Gadea, S., Sedeño, L., Sigman, M., García, A. M., Fraiman, R., & Fraiman, D. (2017). Variability in functional brain networks predicts expertise during action observation. *NeuroImage*, 146, 690–700. <https://doi.org/10.1016/j.neuroimage.2016.09.041>
- Andics, A., & Faragó, T. (2018). Voice perception across species. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 363). Oxford University Press.
- Andics, A., Gácsi, M., Faragó, T., Kis, A., & Miklósi, Á. (2014). Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. *Current Biology*, 24(5), 574–578. <https://doi.org/10.1016/j.cub.2014.01.058>
- Andics, A., McQueen, J. M., & Petersson, K. M. (2013). Mean-based neural coding of voices. *NeuroImage*, 79, 351–360. <https://doi.org/10.1016/j.neuroimage.2013.05.002>
- Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., & Vidnyánszky, Z. (2010). Neural mechanisms for voice recognition. *NeuroImage*, 52(4), 1528–1540. <https://doi.org/10.1016/j.neuroimage.2010.05.048>
- Arnold, B. D., & Wilkinson, G. S. (2011). Individual specific contact calls of pallid bats (*Antrozous pallidus*) attract conspecifics at roosting sites. *Behavioral Ecology and Sociobiology*, 65(8), 1581–1593. <https://doi.org/10.1007/s00265-011-1168-4>
- Aubin, T., Jouventin, P., & Hildebrand, C. (2000). Penguins use the two-voice system to recognize each other. *The Royal Society*, 267, 1081–1087. <https://royalsocietypublishing.org/>
- Bánki, A., Brzozowska, A., Hoehl, S., & Köster, M. (2022). Neural Entrainment vs. Stimulus-Tracking: A Conceptual Challenge for Rhythmic Perceptual Stimulation in Developmental Neuroscience. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.878984>
- Barbero, F. M., Calce, R. P., Talwar, S., Rossion, B., & Collignon, O. (2021). Fast periodic auditory stimulation reveals a robust categorical response to voices in the human brain. *ENeuro*, 8(3). <https://doi.org/10.1523/ENEURO.0471-20.2021>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- 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>
- Bee, M. A., & Gerhardt, H. C. (2002). Individual voice recognition in a territorial frog (*Rana catesbeiana*). *Proceedings of the Royal Society B: Biological Sciences*, 269(1499), 1443–1448. <https://doi.org/10.1098/rspb.2002.2041>
- Belin, P. (2006). Voice processing in human and non-human primates. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 361, Issue 1476, pp. 2091–2107). Royal Society. <https://doi.org/10.1098/rstb.2006.1933>

- Belin, P., Bodin, C., & Aglieri, V. (2018). A “voice patch” system in the primate brain for processing vocal information? In *Hearing Research* (Vol. 366, pp. 65–74). Elsevier B.V. <https://doi.org/10.1016/j.heares.2018.04.010>
- Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker’s voice in right anterior temporal lobe. *NeuroReport*, 14(16), 2809–2812. <https://doi.org/10.1097/01.wnr.00000091689.94870.85>
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403, 309–312.
- Blumenrath, S. H., Dabelsteen, T., & Pedersen, S. B. (2007). Vocal neighbour-mate discrimination in female great tits despite high song similarity. *Animal Behaviour*, 73(5), 789–796. <https://doi.org/10.1016/j.anbehav.2006.07.011>
- Boersma, P. & Weenink D. (2023): Praat: doing phonetics by computer [Computer program]. Version 6.3.17, retrieved 10 September 2023 from <https://www.praat.org>.
- Boros, M., Gábor, A., Szabó, D., Bozsik, A., Gácsi, M., Szalay, F., Faragó, T., & Andics, A. (2020). Repetition enhancement to voice identities in the dog brain. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-60395-7>
- Boros, M., Magyari, L., Morvai, B., Hernández-Pérez, R., & Dror, S. (2024). Neural evidence for referential understanding of object words in dogs. *Current Biology*. <https://doi.org/10.1016/j.cub.2024.02.029>
- Boros, M., Magyari, L., Török, D., Bozsik, A., Deme, A., & Andics, A. (2021). Neural processes underlying statistical learning for speech segmentation in dogs. *Current Biology*, 31(24), 5512–5521.e5. <https://doi.org/10.1016/j.cub.2021.10.017>
- Bourne, G. R., Collins, A. C., Holder, A. M., & McCarthy, C. L. (2001). Vocal Communication and Reproductive Behavior of the Frog *Colostethus beebei* in Guyana. In *Source: Journal of Herpetology* (Vol. 35, Issue 2).
- Briseño-Jaramillo, M., Estrada, A., & Lemasson, A. (2015). Individual voice recognition and an auditory map of neighbours in free-ranging black howler monkeys (*Alouatta pigra*). *Behavioral Ecology and Sociobiology*, 69(1), 13–25. <https://doi.org/10.1007/s00265-014-1813-9>
- Carlson, N. V., Kelly, E. M. K., & Couzin, I. (2020). Individual vocal recognition across taxa: A review of the literature and a look into the future: Individual vocal recognition. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 375, Issue 1802). Royal Society Publishing. <https://doi.org/10.1098/rstb.2019.0479>
- Ceugniet, M., & Izumi, A. (2004). Vocal individual discrimination in Japanese monkeys. *Primates*, 45(2), 119–128. <https://doi.org/10.1007/s10329-003-0067-3>
- Czeibert, K., Baksa, G., Grimm, A., Nagy, S. A., Kubinyi, E., & Petneházy, Ö. (2019). MRI, CT and high resolution macro-anatomical images with cryosectioning of a Beagle brain: Creating the base of a multimodal imaging atlas. *PLoS ONE*, 14(3). <https://doi.org/10.1371/journal.pone.0213458>
- De Rosa, M., Ktori, M., Vidal, Y., Bottini, R., & Crepaldi, D. (2022). Frequency-based neural discrimination in fast periodic visual stimulation. *Cortex*, 148, 193–203. <https://doi.org/10.1016/j.cortex.2022.01.005>

- Deaúx, E. C., Allen, A. P., Clarke, J. A., & Charrier, I. (2016). Concatenation of “alert” and “identity” segments in dingoes’ alarm calls. *Scientific Reports*, 6. <https://doi.org/10.1038/srep30556>
- Delon-Martin, C., Plailly, J., Fonlupt, P., Veyrac, A., & Royet, J. P. (2013). Perfumers’ expertise induces structural reorganization in olfactory brain regions. *NeuroImage*, 68, 55–62. <https://doi.org/10.1016/j.neuroimage.2012.11.044>
- Evans, H. E., & De Lahunta, A. (2012). *Miller's Anatomy of the Dog-E-Book: Miller's Anatomy of the Dog-E-Book*. Elsevier Health Sciences.
- Gábor, A., Andics, A., Miklósi, Á., Czeibert, K., Carreiro, C., & Gácsi, M. (2021). Social relationship-dependent neural response to speech in dogs. *NeuroImage*, 243. <https://doi.org/10.1016/j.neuroimage.2021.118480>
- Gábor, A., Kaszás, N., Faragó, T., Pérez Fraga, P., Lovas, M., & Andics, A. (2022). The acoustic bases of human voice identity processing in dogs. *Animal Cognition*, 25(4), 905–916. <https://doi.org/10.1007/s10071-022-01601-z>
- Gábor, A., Kaszás, N., Miklósi, Á., Faragó, T., & Andics, A. (2019). Interspecific voice discrimination in dogs. *Biologia Futura*, 70(2), 121–127. <https://doi.org/10.1556/019.70.2019.15>
- Gómez-Armenta, J. R., Pérez-Espinosa, H., Fernández-Zepeda, J. A., & Reyes-Meza, V. (2024). Automatic classification of dog barking using deep learning. *Behavioural Processes*, 218, 105028. <https://doi.org/10.1016/j.beproc.2024.105028>
- Hahn, A. H., Hoang, J., McMillan, N., Campbell, K., Congdon, J., & Sturdy, C. B. (2015). Biological salience influences performance and acoustic mechanisms for the discrimination of male and female songs. *Animal Behaviour*, 104, 213–228. <https://doi.org/10.1016/j.anbehav.2015.03.023>
- Hauber, M. E., Woolley, S. M. N., Cassey, P., & Theunissen, F. E. (2013). Experience dependence of neural responses to different classes of male songs in the primary auditory forebrain of female songbirds. *Behavioural Brain Research*, 243(1), 184–190. <https://doi.org/10.1016/j.bbr.2013.01.007>
- Hausfeld, L., De Martino, F., Bonte, M., & Formisano, E. (2012). Pattern analysis of EEG responses to speech and voice: Influence of feature grouping. *NeuroImage*, 59(4), 3641–3651. <https://doi.org/10.1016/j.neuroimage.2011.11.056>
- Hu, L., Zhu, L., Huang, H., Lin, G., Ren, B., & Zhang, J. (2021). Speaker recognition with voice evoked EEG. *Proceedings - 2021 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021*, 2231–2236. <https://doi.org/10.1109/BIBM52615.2021.9669324>
- Janik, V. M., & Sayigh, L. S. (2013). Communication in bottlenose dolphins: 50 years of signature whistle research. In *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* (Vol. 199, Issue 6, pp. 479–489). <https://doi.org/10.1007/s00359-013-0817-7>
- Jia, G., Bai, S., Lin, Y., Wang, X., Zhu, L., Lyu, C., Sun, G., An, K., Roe, A. W., Li, X., & Gao, L. (2023). Special Topic: Non-human Primates as Animal Model Systems Representation of conspecific vocalizations in amygdala of awake marmosets. *National Science Review*, 10(11). <https://doi.org/10.1093/nsr/nwad194>

- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002–1011.
<https://doi.org/10.1111/j.1467-7687.2011.01052.x>
- Kanber, E., Lavan, N., & Mcgettigan, C. (2022). Highly accurate and robust identity perception from personally familiar voices. *J Exp Psychol Gen.*, 151(4), 897–911.
<https://doi.org/10.1037/xge0001112>
- Kelley, D. B. (2004). Vocal communication in frogs. In *Current Opinion in Neurobiology* (Vol. 14, Issue 6, pp. 751–757). <https://doi.org/10.1016/j.conb.2004.10.015>
- Kirk, U., Skov, M., Christensen, M. S., & Nygaard, N. (2009). Brain correlates of aesthetic expertise: A parametric fMRI study. *Brain and Cognition*, 69(2), 306–315.
<https://doi.org/10.1016/j.bandc.2008.08.004>
- Kis, A., Szakadát, S., Kovács, E., Gácsi, M., Simor, P., Gombos, F., Topál, J., Miklósi, Á., & Bódizs, R. (2014). Development of a non-invasive polysomnography technique for dogs (*Canis familiaris*). *Physiology and Behavior*, 130, 149–156.
<https://doi.org/10.1016/j.physbeh.2014.04.004>
- Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, K., & Wang, Z. (2003). Effects of Experience on Fetal Voice Recognition. *Psychological Science*, 14(3), 220–224.
- Klaumann, P. R., Moreno, J. C. D., & Montiani-Ferreira, F. (2018). A morphometric study of the canine skull and periorbita and its implications for regional ocular anesthesia. *Veterinary Ophthalmology*, 21(1), 19–26. <https://doi.org/10.1111/vop.12471>
- Kojima, S., Izumi, A., & Ceugniet, M. (2003). Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates*, 44(3), 225–230.
<https://doi.org/10.1007/s10329-002-0014-8>
- 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>
- Koyano, K. W., Jones, A. P., McMahon, D. B. T., Waidmann, E. N., Russ, B. E., & Leopold, D. A. (2021). Dynamic Suppression of Average Facial Structure Shapes Neural Tuning in Three Macaque Face Patches. *Current Biology*, 31(1), 1–12.e5.
<https://doi.org/10.1016/j.cub.2020.09.070>
- Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, 22(2), 948–955.
<https://doi.org/10.1016/j.neuroimage.2004.02.020>
- Kriengwatana, B., Escudero, P., & Cate, C. ten. (2014). Revisiting vocal perception in non-human animals: A review of vowel discrimination, speaker voice recognition, and speaker normalization. In *Frontiers in Psychology* (Vol. 5, Issue OCT). Frontiers Media S.A.
<https://doi.org/10.3389/fpsyg.2014.01543>
- Lakatos, P., Gross, J., & Thut, G. (2019). A New Unifying Account of the Roles of Neuronal Entrainment. In *Current Biology* (Vol. 29, Issue 18, pp. R890–R905). Cell Press.
<https://doi.org/10.1016/j.cub.2019.07.075>
- Larrañaga, A., Bielza, C., Pongrácz, P., Faragó, T., Bálint, A., & Larrañaga, P. (2015). Comparing supervised learning methods for classifying sex, age, context and individual Mudi dogs

- from barking. *Animal Cognition*, 18(2), 405–421. <https://doi.org/10.1007/s10071-014-0811-7>
- Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2(JUL), 1–12. <https://doi.org/10.3389/fpsyg.2011.00175>
- Latinus, M., McAleer, P., Bestelmeyer, P. E. G., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, 23(12), 1075–1080. <https://doi.org/10.1016/j.cub.2013.04.055>
- Latinus, M., & Taylor, M. J. (2012). Discriminating male and female voices: Differentiating pitch and gender. *Brain Topography*, 25(2), 194–204. <https://doi.org/10.1007/s10548-011-0207-9>
- Lavan, N., Knight, S., & McGettigan, C. (2019). Listeners form average-based representations of individual voice identities. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-10295-w>
- Lavan, N., Kreitewolf, J., Obleser, J., & McGettigan, C. (2021). Familiarity and task context shape the use of acoustic information in voice identity perception. *Cognition*, 215. <https://doi.org/10.1016/j.cognition.2021.104780>
- Lehoczki, F., Szenczi, P., Bánszegi, O., Lakatos, K., & Faragó, T. (2020). Cross-species effect of separation calls: family dogs' reactions to pup, baby, kitten and artificial sounds. *Animal Behaviour*, 168, 169–185. <https://doi.org/10.1016/j.anbehav.2020.08.015>
- Lind, H., Dabelsteen, T., & McGregor, P. K. (1996). Female great tits can identify mates by song. *Animal Behaviour*, 52, 667–671.
- Luck, S. J. (2014). An introduction to the event-related potential technique. MIT Press.
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, 97(8), 4398–4403. <https://doi.org/10.1073/pnas.070039597>
- Magyari, L., Huszár, Z., Turzó, A., & Andics, A. (2020). Event-related potentials reveal limited readiness to access phonetic details during word processing in dogs: Word processing in dogs. *Royal Society Open Science*, 7(12). <https://doi.org/10.1098/rsos.200851>
- McComb, K., Reby, D., Baker, L., Moss, C., & Sayialel, S. (2003). Long-distance communication of acoustic cues to social identity in African elephants. *Animal Behaviour*, 65(2), 317–329. <https://doi.org/10.1006/anbe.2003.2047>
- McDonald, P. G. (2012). Cooperative bird differentiates between the calls of different individuals, even when vocalizations were from completely unfamiliar individuals. *Biology Letters*, 8(3), 365–368. <https://doi.org/10.1098/rsbl.2011.1118>
- Mehler, J., Bertoncini, J., Barriere, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. In *Perception* (Vol. 7).
- Miklósi, Á. (2015). *Dog behaviour, evolution and cognition*.
- Molnár, C., Kaplan, F., Roy, P., Pachet, F., Pongrácz, P., Dóka, A., & Miklósi, Á. (2008). Classification of dog barks: A machine learning approach. *Animal Cognition*, 11(3), 389–400. <https://doi.org/10.1007/s10071-007-0129-9>

- Molnár, C., Pongrácz, P., Dóka, A., & Miklósi, Á. (2006). Can humans discriminate between dogs on the base of the acoustic parameters of barks? *Behavioural Processes*, 73(1), 76–83. <https://doi.org/10.1016/j.beproc.2006.03.014>
- Molnár, C., Pongrácz, P., Faragó, T., Dóka, A., & Miklósi, Á. (2009). Dogs discriminate between barks: The effect of context and identity of the caller. *Behavioural Processes*, 82(2), 198–201. <https://doi.org/10.1016/j.beproc.2009.06.011>
- Mongelli, V., Dehaene, S., Vinckier, F., Peretz, I., Bartolomeo, P., & Cohen, L. (2017). Music and words in the visual cortex: The impact of musical expertise. *Cortex*, 86, 260–274. <https://doi.org/10.1016/j.cortex.2016.05.016>
- Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottareau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, 15(6), 1–46. <https://doi.org/10.1167/15.6.4>
- Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. In *Trends in Cognitive Sciences* (Vol. 23, Issue 11, pp. 913–926). Elsevier Ltd. <https://doi.org/10.1016/j.tics.2019.08.004>
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, JM (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, Volume 2011 (2011), Article ID 156869, doi:10.1155/2011/156869
- Palacios, V., Font, E., Márquez, R., & Carazo, P. (2015). Recognition of familiarity on the basis of howls: A playback experiment in a captive group of wolves. *Behaviour*, 152(5), 593–614. <https://doi.org/10.1163/1568539X-00003244>
- Perrodin, C., & Petkov, C. I. (2018). Voice-Sensitive Regions, Neurons, and Multisensory Pathways in the Primate Brain. In S. Frühholz & P. Belin (Eds.), *The Oxford Handbook of Voice Perception* (pp. 337). Oxford University Press.
- Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience*, 11(3), 367–374. <https://doi.org/10.1038/nn2043>
- Petrinovich, L. (1974). Individual Recognition of Pup Vocalization by Northern Elephant Seal Mothers. *Zeitschrift Für Tierpsychologie*, 34(3), 308–312. <https://doi.org/10.1111/j.1439-0310.1974.tb01803.x>
- Pinaud, R., & Terleph, T. A. (2008). A songbird forebrain area potentially involved in auditory discrimination and memory formation. *Journal of Biosciences*, 33, 145–155. <http://www.ias.ac.in/jbiosci>
- Plante-Hébert, J., Boucher, V. J., & Jemel, B. (2021). The processing of intimately familiar and unfamiliar voices: Specific neural responses of speaker recognition and identification. *PLoS ONE*, 16(4 April). <https://doi.org/10.1371/journal.pone.0250214>
- Pongrácz, P., Molnár, C., Miklósi, Á., & Csányi, V. (2005). Human listeners are able to classify dog (*Canis familiaris*) barks recorded in different situations. *Journal of Comparative Psychology*, 119(2), 136–144. <https://doi.org/10.1037/0735-7036.119.2.136>
- Portfors, C. V., Roberts, P. D., & Jonson, K. (2009). Over-representation of species-specific vocalizations in the awake mouse inferior colliculus. *Neuroscience*, 162(2), 486–500. <https://doi.org/10.1016/j.neuroscience.2009.04.056>

- R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Ratcliffe, V. F., McComb, K., & Reby, D. (2014). Cross-modal discrimination of human gender by domestic dogs. *Animal Behaviour*, 91, 127–135. <https://doi.org/10.1016/j.anbehav.2014.03.009>
- Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. In *Anim. Behav* (Vol. 51).
- Retter, T. L., & Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia*, 91, 9–28. <https://doi.org/10.1016/j.neuropsychologia.2016.07.028>
- Root-Gutteridge, H., Ratcliffe, V. F., Korzeniowska, A. T., & Reby, D. (2019). Dogs perceive and spontaneously normalize formant-related speaker and vowel differences in human speech sounds. *Biology Letters*, 15(12). <https://doi.org/10.1098/rsbl.2019.0555>
- Ruthig, P., & Schönwiesner, M. (2022). Common principles in the lateralization of auditory cortex structure and function for vocal communication in primates and rodents. In *European Journal of Neuroscience* (Vol. 55, Issue 3, pp. 827–845). John Wiley and Sons Inc. <https://doi.org/10.1111/ejn.15590>
- Stevenage, S. V., Symons, A. E., Fletcher, A., & Coen, C. (2020). Sorting through the impact of familiarity when processing vocal identity: Results from a voice sorting task. *Quarterly Journal of Experimental Psychology*, 73(4), 519–536. <https://doi.org/10.1177/1747021819888064>
- Šuta, D., Kvašňák, E., Popelář, J., & Syka, J. (2003). Representation of Species-Specific Vocalizations in the Inferior Colliculus of the Guinea Pig. *Journal of Neurophysiology*, 90(6), 3794–3808. <https://doi.org/10.1152/jn.01175.2002>
- Šuta, D., Popelář, J., Burianová, J., & Syka, J. (2013). Cortical Representation of Species-Specific Vocalizations in Guinea Pig. *PLoS ONE*, 8(6). <https://doi.org/10.1371/journal.pone.0065432>
- Šuta, D., Popelář, J., Kvašňák, E., & Syka, J. (2007). Representation of species-specific vocalizations in the medial geniculate body of the guinea pig. *Experimental Brain Research*, 183(3), 377–388. <https://doi.org/10.1007/s00221-007-1056-3>
- Syka, J., Šuta, D., & Popelář, J. (2005). Responses to species-specific vocalizations in the auditory cortex of awake and anesthetized guinea pigs. *Hearing Research*, 206(1–2), 177–184. <https://doi.org/10.1016/j.heares.2005.01.013>
- Szipl, G., Boeckle, M., Wascher, C. A. F., Spreafico, M., & Bugnyar, T. (2015). With whom to dine? Ravens' responses to food-associated calls depend on individual characteristics of the caller. *Animal Behaviour*, 99, 33–42. <https://doi.org/10.1016/j.anbehav.2014.10.015>
- Tagliabata, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2009). Visualizing vocal perception in the chimpanzee brain. *Cerebral Cortex*, 19(5), 1151–1157. <https://doi.org/10.1093/cercor/bhn157>
- Taylor, A. M., Ratcliffe, V. F., McComb, K., & Reby, D. (2014). Auditory Communication in Domestic Dogs: Vocal Signalling in the Extended Social Environment of a Companion Animal. In *The Social Dog: Behavior and Cognition* (pp. 131–163). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-407818-5.00005-X>

- Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., & Sawyer, S. K. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science*, 342(November), 871–874.
- Vos, S., Collignon, O., & Boets, B. (2023). The Sound of Emotion: Pinpointing Emotional Voice Processing Via Frequency Tagging EEG. *Brain Sciences*, 13(2). <https://doi.org/10.3390/brainsci13020162>
- 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>
- Weary, D. M., & Krebs, J. R. (1992). Great tits classify songs by individual voice characteristics. *Animal Behaviour*, 43, 283–287.
- Weeden, J. S., & Falls, J. B. (1959). Differential responses of male ovenbirds to recorded songs of neighboring and more distant individuals. *Auk*, 76, 343–351.
- Weston, P. S. J., Hunter, M. D., Sokhi, D. S., Wilkinson, I. D., & Woodruff, P. W. R. (2015). Discrimination of voice gender in the human auditory cortex. *NeuroImage*, 105, 208–214. <https://doi.org/10.1016/j.neuroimage.2014.10.056>
- Wooller, R. D. (1978). Individual Vocal Recognition in the Kittiwake Gull, *Rissa tridactyla* (L.). *Ethology*, 48(1), 68–86. <https://doi.org/10.1111/j.1439-0310.1978.tb00248.x>
- Zäske, R., Volberg, G., Kovács, G., & Schweinberger, S. R. (2014). Electrophysiological correlates of voice learning and recognition. *Journal of Neuroscience*, 34(33), 10821–10831. <https://doi.org/10.1523/JNEUROSCI.0581-14.2014>

Acknowledgements

First of all, I would like to thank my supervisors, Boglárka Morvai and Attila Andics, who spared no time and energy to help me in my work and improvement. Thank you Bogi for guiding me through this whole experiment, taking my scripting skills and statistical knowledge to a higher level and challenging my critical thinking. I am grateful for your patience and that I could always have another question. I am very lucky to have had the chance to work with you, learn from you and get to know you better over the past year. Attila, thank you for taking me and this project seriously. I still do not always understand why, but you usually believe in me more than I believe in myself. Sometimes I fail to express properly but I am very grateful for all your support, really.

Special thanks to the human subjects for their time and willingness to participate in the study. I am grateful to the owners for their enthusiasm, curiosity and for bringing their animals to the Department. Most importantly, I would like to express my deepest gratitude to all the dogs for their cooperation, patience, endurance, and calmness during the measurements. They did an amazing job of letting us decorate them with electrodes and staying calm for several minutes while listening to our crazy sound streams. Last but not least, I would like to thank Kitti Szabó and Franceska Furik for their contribution to the data collection, but most of all for all the support they give me every day.

2024.05.08.