Studying speech sound discrimination in dogs

Philomena Athanasiadou

Studying speech sound discrimination in dogs

Philomena Athanasiadou

philo.ath@gmail.com

Thesis, 60 ECTS

Master in Cognitive Artificial Intelligence,

Utrecht University

January 2012

Review committee:

dr. Annemarie Kerkhoff (first supervisor)

prof. dr. Frank Wijnen

dr. Ignace Hooge

Abstract

Speech sound perception has readily been studied in non-human primates and rats illustrating that we share some of the capabilities underlying speech perception with other animals. However an animal with great potential has been overlooked. Dogs (*Canis lupus familiaris*) share an intimate history with humans that started thousands of years ago. The process of domestication has made dogs highly attuned to our vocalizations and body language and their reputation for being 'man's best friend' is surely not unfounded. In this thesis, I argue that there are great benefits to be obtained from studying speech sound perception in dogs. I suggest using the preferential looking paradigm to do so, which is a method widely used for human infant studies and therefore allows for immediate comparison of results. By conducting a simple experiment of speech sound discrimination, I aim to answer: (1) whether the preferential looking paradigm can be used to study speech sound discrimination in dogs; and (2) whether dogs that have grown up in Dutch-speaking homes can discriminate between the Dutch vowels /a/ and /e/ (embedded in the non-words *faap* and *feep*). The answer to both questions is positive, thus opening the way for future research.

Keywords: speech sound perception; phoneme discrimination; preferential looking; dogs

Acknowledgments

First of all I want to thank Annemarie Kerkhoff and Frank Wijnen for being the open-minded, kind, and inspiring people that they are. They introduced me to psycholinguistics and were not scared off when I arrived with my own research proposal about "canine linguistics". Instead they agreed to embark with me on this adventure and enthusiastically supported me.

Of course, none of this would have been possible if it had not been for the dog daycare centers and all the dog owners that agreed to participate in this study. I heartily thank them and especially: Olga van den Engel, Amy Looijen, Kees & Anja and Karin Spies. Their willingness to participate, trust me, and lend me some of their time is what made this research possible.

I thank my grandfather, for generously giving me his garage and all sorts of materials, as well as his time and know-how for building the experimental set-up. (Grandma, thanks for nursing me with good food during this time and making sure I take some breaks ;)).

I also want to thank the following people for taking the time to read through material related to this thesis (or even the whole thesis itself!), and offer their knowledge, comments and ideas for improvement: Sibrecht Bouwstra, Desiree Capel, Ignace Hooge, Maria Stratemeier, Liquan Liu, Maartje de Klerk, Marjolein van Egmond, Luca Valerio and Karin Hendrick.

My thanks also go to: Steven Duursma for lending his webcam for this study and printing the "certificates of participation"; Pierre-Antoine Benoist for his advice on various occasions and for patiently coding the video-recordings; Theo Veenker for taking the time to introduce me to ZEP and share his view on the experimental procedure; Sjef Pieters for advising me on the proper equipment; Matthijs Schilder and Claudia Vinke for sharing their expertise on studies with dogs; Fred Poelma for advising me on the ethical aspects of my study; and Xavier, Katerina and Maartje for being my Guinea-pigs while preparing for the experiments.

Finally, to all the cute, sweet, funny, wonderful dogs that participated in this study I say: "Woof, woof, woof!"

Disclaimer:

No dogs were hurt during this study.

Permissions were obtained from the owners for displays their dogs' images in this thesis.

Contents

(1/2)

Contents

(2/2)

1. Introduction

One of the most thorny subjects concerning artificial intelligence is the role of language. Do thoughts require language? Does intelligent behavior require a symbolic abstraction of reality? Language, our sophisticated communication medium, seems to lie at the very foundation of human culture and achievements. But how unique are human linguistic abilities really? This is the question that most fascinates me and that has led to this thesis.

Abilities underlying the acquisition and comprehension of spoken language have long been assumed to be 'language specific' and therefore uniquely human. Over the past few years however, researchers have shown that a lot of these abilities are neither 'language specific', nor uniquely human. Naturally, given our genetic relatedness, the abilities of non-human primates were the first to be investigated. Soon non-human primates demonstrated their ability to segment speech, discriminate between languages and even learn some types of non-adjacent and predictive dependencies between words (Hauser et al. 2001, Ramus et al. 2000, Newport & Aslin 2004, Saffran et al. 2008). More recently also rodents and birds have been studied. Chinchillas can reliably distinguish English phonemes (Burdick & Miller 1975, Kuhl & Miller 1975). Rats can segment speech and tell Japanese and Dutch apart (Toro & Trobalon 2005; Toro et al. 2005). A bird, the European starling, can learn intricate sound patterns and distinguish "grammatical" from "ungrammatical" streams (Gentner et al. 2006).

The above mentioned studies focus on categorical perception (the ability to group speech sounds into phonemes) and statistical learning (the ability to track co-occurrence and interdependence of phonemes, syllables and words). Unfortunately, none of these abilities have ever been studied systematically in dogs. As a result, the only sources we can use to study dogs' "linguistic abilities" are popular beliefs, anecdotes and studies whose primary focus lies somewhere else (e.g. obedience, the role of the auditory cortex in speech sound perception and dogs' ability to learn object names). Given that dogs have had an intimate relationship with humans for thousands of years, are exposed to language on a daily basis, readily respond to our vocalizations and are everywhere (i.e. no need to raise them in labs), I expect that great benefits can be gained from studying speech sound perception in dogs.

The ability to discriminate between speech sounds lies at the heart of all the aforementioned studies (i.e. discriminating between phonemes, sequences, languages). Therefore, in this thesis, I propose studying speech sound discrimination in dogs using the preferential looking paradigm. This method is non-invasive, requires no training and has widely been used for studying speech sound discrimination in human infants.

2. Why dogs?

2.1. The dog: a unique animal

Dogs and humans share a long and intimate history. Dogs evolved from their wolf-like ancestors through the process of domestication. There are different views on the exact nature of this process, but they all agree on this: contact with humans made dogs what they are today. Based on genetic evidence, this process is estimated to have started as early as 100,000 years ago (Vilá et al. 1997). The earliest archeological evidence consists of dog remains found in human graves dating back to 13,000 BP (Nobis 1979, as cited in Miklósi 2007). One of the most remarkable findings is the skeleton of a puppy found in a human grave from 11,000 BP (Davis & Valla 1978). The puppy was buried next to the human with the human's hand placed over its body. These findings would seem to suggest that not only did dogs and humans live in proximity and share the same environment, they also formed affectionate relationships.

It is important to realize at this point that dogs are not mere tame wolves. Studies in which wolf cubs were hand-raised by humans and intensively socialized show that differences remain. Dogs can readily understand and interpret humans' pointing gestures at the age of 2–4 months while human-raised wolves are only able to do this at a much later age (>1.5 years) (Miklósi et al. 2003). Furthermore, dogs often search eye contact with humans while wolves do not (Miklósi et al 2003, Virányi et al 2008). Wolves do not exploit this human communication cue and might even perceive it as a threat (Vas et al 2005). When faced with an impossible task for example, dogs seek help from a human by looking at them while the socialized wolves never turn towards a human (Miklósi et al. 2003). Finally, puppies show a pattern of attachment behavior towards their human caretaker while wolf cubs do not (Kubinyi et al 2007). The process of domestication seems to have fundamentally changed the dog species. A study that tried to replicate this process with foxes illustrates this. Foxes, just like wolves, have no natural affinity for humans. By selecting foxes on their "tameness", researchers ended up with a type of fox that not only resembled dogs in their behavior but also in their physique (Belyaev 1979). These foxes were more interested in humans and attuned to their communicative cues, vocalized and whimpered more and wagged their tails. They also started to grow curly tails, droopy ears and sometimes had a patched or multicolored fur (features absent in wild foxes).

Several researchers argue that, after thousands of years of sharing the same ecological and social environment with humans, dogs have developed unique social and communicative abilities (Miklósi et al. 2004, Hare and Tomasello 2005, Topál et al. 2009). This, they argue, makes dogs very interesting for comparative studies. Bloom (2004) in fact says that dogs might just become the "new chimpanzees" to psychologists.

2.2. What we can infer about dogs' 'linguistic abilities'

Though we train dogs to respond to our vocalizations, no studies have been done to directly study the abilities underlying language comprehension and acquisition (such as categorical perception and statistical learning) in dogs, as done with other animals. There are however some studies that involve speech sound perception in dogs. One of the earliest studies (Buytendijk & Fischer 1936, as cited in Miklósi 2007) tested whether dogs respond reliably to commands and what the effects of systematically changing the phonemes of these commands are. Buytendijk & Fischer observed that dogs were more likely to respond to the commands when the changes were made at the end of the word instead of at the beginning. Baru & Shmigidina (1977) tested for synthesized vowel discrimination in dogs ([a] vs [i]) before and after surgical removal of the auditory cortex. The intact dogs could discriminate between the two vowels irrespective of stimulus intensity, voice frequency and duration. This should not come as a surprise as it confirms the observation people make in everyday life while interacting with dogs: dogs tend to respond to their names irrespective of the person calling them.

More recently, Kaminski et al. (2004) reported that Rico, a Border Collie, knew the names of 200 objects, and could correctly retrieve them upon request. Furthermore, Rico also proved capable of inferring the name of novel objects through exclusion and remembered the object's names until 4 weeks later (this process is called 'fast-mapping' and was initially thought to be uniquely human). Pilley and Reid (2011) were able to teach their Border Collie, Chaser, the names of 1,022 objects through an intensive 3-year long training. Chaser could combine the object names with any of three commands ('take', 'paw' and 'nose'), illustrating that she was able to distinguish the individual words and that she was not reacting to a single stream of speech. Just like Rico, Chaser also learned by exclusion, but most impressive is her ability to abstract to object categories. She was able to learn three common nouns ('toy', 'ball' and 'Frisbee') representing three categories of objects (with 'ball' and 'Frisbee' being subsets of 'toy'). When asked to fetch an object from a category (i. e. 'fetch a ball'), Chaser would correctly retrieve an object from that category (note that these objects differed in color and size). Chaser was also able to do this with objects different from the ones used during training. This shows that Chaser can generalize (probably using the common physical properties of the objects) and apply category labels to novel objects. Interestingly, she also understood the difference between ‘toys’ (that she was allowed to play with) and ‘non-toys’, that were very similar to toys in terms of shape and size but belonged to other members of the household.

Finding that dogs can learn up to 1,022 words, understand object categories and learn by exclusions is surprising. One of the questions remaining is whether any dog, given the proper training, is capable of such feats. It might be possible that these two Border Collies are exceptions, or that their demonstrated skills are breed-specific. Still, given the above, I suspect that the average dog must not have developed some interesting strategies for processing human speech that are worth investigating.

3. Research questions

This study has two aims: (1) To propose the preferential looking paradigm for testing speech sound discrimination in dogs; (2) To apply this method to validate whether dogs are sensitive to the difference between /a/ and /e/ (two Dutch vowels).

3.1. The preferential looking paradigm

As mentioned previously, the ability to discriminate speech sounds lies at the heart of research that studies the abilities underlying language comprehension and acquisition (such as categorical perception and statistical learning) in animals and human infants. So how can we test speech sound discrimination in dogs? For testing whether dogs can hear the difference between the synthesized vowels [a] and [i], Baru & Shmigidina (1977) taught the dogs to raise either their left or their right paw depending on the vowel. However, this method has the following two drawbacks. First of all, it is time consuming as dogs need to be trained and tested on their obedience beforehand. Secondly, as Howell et al. (2011) point out, obedience and motivation can be confounding factors in this type of research. Here a failure in discrimination is not necessarily a failure in perception. It might just be a failure in motivation or obedience. Clearly, a method without such shortcomings would be preferable.

A method used in a wide range of infant studies is that of preferential looking. The underlying idea is the following: imagine that you want to test whether infants perceive a difference between stimulus A and stimulus B. First the infant is exposed to stimulus A until the infant loses interest and stops attending to the stimulus (this is called the familiarization or habituation phase). Then stimulus B is presented. If the infant recovers interest (the infant starts attending to stimulus B, e.g. looking at it), one can conclude that he/she perceives a difference. On the other hand, if stimulus A continued to be presented instead, the infant is expected to remain uninterested and hardly attend to this stimulus. Of course in reality such extremes as ‘attending to the stimulus all the time’ and ‘not attending at all’ are rarely encountered. Using the infant’s gaze as an indicator of interest, the time the infant spends looking at each stimulus is measured (the looking time) and compared. It is even common practice to use the average looking time from multiple trials. If the time the infant spends looking at stimulus A is significantly different from time spent looking at stimulus B, we may conclude that the infant perceives a difference between the two stimuli.

When testing for discrimination of auditory stimuli, there is nothing to 'look at' technically speaking (except for looking in the direction of the sound source). Therefore auditory stimuli are often accompanied by visual stimuli in these experiments (like images or videos of human faces). Given that this method is widely accepted and used in infant speech perception studies, I propose to use this method with dogs.

Even though the preferential looking paradigm has not been used in speech perception studies with animals (such as the cotton-top tamarins and rats mentioned in section 1), it has been used successfully with dogs to study other (language unrelated) abilities. Researchers used looking time to study whether dogs can discriminate between the correct and incorrect outcome of an arithmetic operation (West & Young, 2002), whether they can match their owner's voice to his or her picture (Adachi et al., 2007), whether they can match another dog's growl to its body size (Faragó et al., 2010) and whether they can detect an unnatural change in an object's size (Müller et al. 2011). However, in all of the above experiments visual stimuli or a combination of visual and auditory stimuli were used. This means that the method should be somewhat adapted in order for it to work for auditory stimuli only. The most obvious solution is to accompany the auditory stimuli by visual stimuli, such as images or videos, as is done in infant studies. Another solution would be to put a salient object right in front of the sound source as a point of visual fixation for the dog. Given that the first solution requires sophisticated equipment (like a LCD monitor) that did not fit my research budget and is impractical to carry around, I opted for the second solution.

3.2. Native versus non-native contrasts

One of the most basic acts of speech sound discrimination would be to discriminate between two phonemes. Currently, there is research being conducted at the Babylab in the Utrecht Instititute of Linguistics that studies infants' ability to discriminate the native (Dutch) contrast /a/ and /e/, and the non-native contrast /æ/ and /ɛ/. I decided to use the same material and start by testing dogs' discrimination abilities on the /a/ and /e/ contrast embedded in the words *faap* and *feep* (as used by de Bree et al. 2010). I expect this to be within dogs' capacities since dogs can discriminate between different commands and respond accordingly. Furthermore, the dog Chaser knew 1,022 object names (Pilley and Reid 2011), and Baru & Shmigidina (1977) showed that dogs can discriminate between the synthesized vowels [a] and [i] irrespective of stimulus intensity, voice frequency, and duration. By testing discrimination on something that we consider to be within the dogs' capacities, we can focus on the method used and assess its validity.

Studying dogs' abilities to discriminate /a/ and /e/ (a native Dutch contrast) has an additional interest. To explain this, I will make a small digression into language acquisition. At birth, human infants are able to hear the difference between phonetic contrasts in any natural language. As they grow older, starting at 6 months of age (Kuhl 2004), they start grouping (categorizing) these sounds and learn to focus only on those phonetic contrast that are meaningful in their native language. As the research of de Bree et al. (2010) shows, 6 and 8 month old Dutch infants hear the difference between /a/ and /e/ (a Dutch native contrast), while they do not hear the difference between /æ/ and /ɛ/ (which is a non-native contrast).

While infants naturally acquire this 'categorical' perception of speech, they are not the only ones to do so. Several non-human primate, rodent and bird species show the same abilities after extensive exposure to language in a lab setting (Hauser 1996, Ramus et al. 2000, Kuhl & Padden 1983, Kuhl & Miller 1975). Since dogs are exposed to language on a daily basis, I expect them to also perceive speech categorically. Thus, after having validated that dogs can indeed hear the difference between /a/ and /e/, one could test them on the non-native contrast /æ/ and /ɛ/. If it turns out that dogs can distinguish the native contrast but not the non-native contrast, then dogs will be the first species shown to readily and naturally (i.e. without prolonged exposure in a lab to a limited set of stimuli) categorize speech sounds from the ambient language.

4. Pilot experiment

The goal of this pilot study was twofold. The first goal was to assess whether the chosen method and set-up is appropriate for dogs. The second goal was to assess whether dogs can discriminate between the non-words *feep* and *faap* (i.e. hear the difference between /e/ and /a/ which is a Dutch native contrast). To do this I adapted the “hybrid” variant of the visual habituation procedure (VHP) as described by Houston et al. (2007), which is a method used for assessing speech sound discrimination in individual infants. The paradigm consists of a habituation phase and a test phase. During the habituation phase the dog is habituated to the non-word *feep.* Subsequently he/she is exposed to two types of test trials. On *old* trials the dog hears repetitions of the same non-word he/she heard during habituation (i.e. “*feep, feep, feep, feep...*”). On *new* trials a *feep* token alternates with a token of the non-word not heard during familiarization, namely *faap* (i.e. *“feep, faap, feep, faap...”*). The expectation is that dogs (just as infants) will show longer looking times during *new* trials than during *old* trials.

4.1. Method

4.1.1. Participants

Dogs from Dutch-speaking homes were recruited through three dog daycare centers (which also served as the test locations). Owners who had interest in the study and whose dog would be present on the day of testing, signed a consent form that allowed their dog to participate. Three dogs from various breeds participated in this study: two male, aged 10 years, 9 months and 9 years, 5 months respectively, and one female of unknown age. None of the dogs had earlier experience with similar experiments.

4.1.2. Stimuli

The stimuli used consist of five tokens of the monosyllabic non-words *feep* (voice no. 1, 2, 3, 4 and 5) and one token of the word *faap* (voice no. 5). These tokens were recorded by 5 female staff members of the Babylab Utrecht, and had a mean duration of 620 ms. These are the same stimuli there were used in a study on infants’ ability to discriminate native versus non-native contrasts (de Bree et al. 2010).

4.1.3. Apparatus

The test was conducted at three dog daycare centers across the Netherlands with a portable set-up. The set-up was assembled on location, in the most isolated/noise free room of each establishment. Dogs sat at the feet of their human companion (their owner or a daycare staff member), facing a wooden frame of 40 cm by 60 cm. The wooden frame was located at a distance of approximately 2 meters from the dog. Behind the frame two loudspeakers were hidden through which the auditory stimuli were presented. A wooden puppet, whose limbs could be moved by pulling a string, was hung in front of the wooden frame and used as an attention getter. Behind the frame a beach-tent hid the experimenter (myself). I operated the wooden puppet by pulling the string and controlled the presentation of the auditory stimuli through the "experiment laptop". On the "experiment laptop" run software written by myself in ZEP (Veenker 2011). During the experiment I observed the dog's reactions in real time through a camera that was mounted on the wooden frame and connected to another laptop, the "observation laptop" (which also recorded this input). Additionally, a second camera recorded the complete scene from the side and was used during analysis for verification. For a better understanding of the experimental set-up see Figures 1 to 5.

Finally, the dogs' owners filled in an online questionnaire. The questionnaire was inspired by C-BARQ[1] and online questionnaires for dog participants of the Universities of Budapest[2], Vienna[3] and Florida[4]. The aim of the questionnaire was: (a) to gather basic information about the dog such as age, gender, breed and household composition; (b) to screen for health and behavioral problems; (c) to get an indication of the quantity and quality of contact the dog has with humans, his/her language exposure and lexicon; (d) to measure the dog's personality and the extent to which he/she is 'human oriented'; and (e) to get some basic information about the owner (see Appendix A). This information was gathered as a basis for future studies, as well as to examine whether some measures might correlate with the dogs ability to discriminate the words.

[1] C-BARQ (Canine Behavioral Assessment and Research Questionnaire): http://vetapps.vet.upenn.edu/cbarq/
[2] Family Dog project, department of ethology, Eötvös Loránd University, Budapest. Participant questionnaire: https://spreadsheets.google.com/viewform?formkey=dHhXUGQ4UnNoVzd1MTVxM1lfamFNU2c6MQ
[3] Clever Dog Lab, department of cognitive biology, University of Vienna, Vienna. Participant questionnaire: https://spreadsheets.google.com/viewform?formkey=dHBMSExrb0VZTjNKRXdsWkVjeWotZ2c6MQ
[4] Canine Cognition and Behavior Lab, department of Psychology, University of Florida, Gainesville. Participant questionnaire: https://spreadsheets.google.com/viewform?key=pLZwuC_igZKNVrcDucggFIg

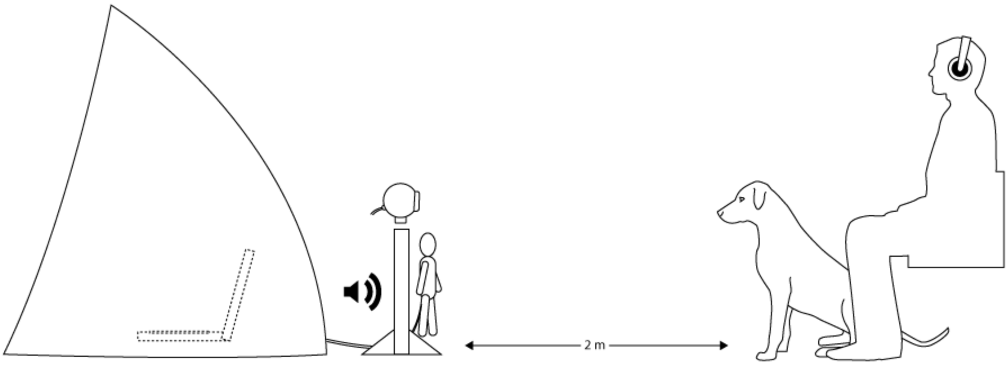

Figure 1. Overview of the experimental set-up.

Figure 2. Participant's view of the experimental set-up.

Figure 3. Researcher's view of the experimental set-up.

Figure 4. Front view camera.

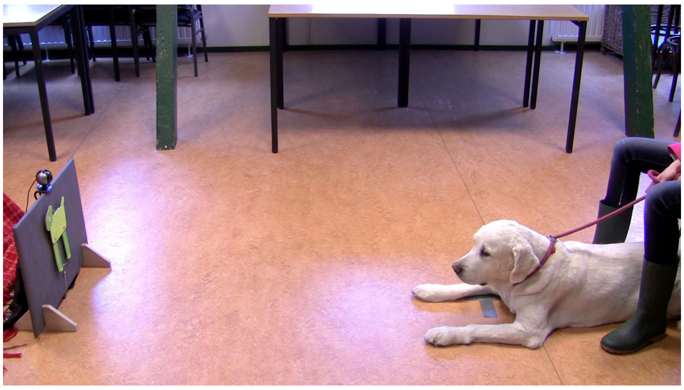

Figure 5. Side view camera.

4.1.4. Procedure

The participating dog entered the experiment room accompanied by his/her human companion. The dog was allowed to explore the room and sniff around for about 2 minutes, after which the companion was seated on a chair with the dog in front of him/her. The companion remained present throughout the test and lightly restrained the dog. Companions were instructed to ignore the test and the dog's behavior as much as possible and only intervene (non-verbally) when the dog ventured too far off from his/her original seating location or turned his/her back towards the loudspeakers for too long. To further limit the possibilities of (involuntary) human cueing the companions wore headphones during the experiment which played a mix of masking rain, rainforest and flowing water sounds.

The experiment was based on the hybrid variant of the Visual Habituation Procedure (VHP; Houston et al. 2007). It started with the habituation phase which was promptly followed by the test phase. Hidden behind a beach-tent I observed the dog's behavior and pressed keys on the 'experiment laptop' in order to indicate whether the dog was looking towards or facing away from the loudspeakers. These keypresses controlled the stimulus presentation and were registered by the experimental ZEP program.

The trials. Before each trial, the puppet's string was pulled repeatedly and in a steady and slow rhythm until the dog oriented towards it (the loudspeakers were hidden just behind the puppet). Once the dog looked in the direction of the loudspeakers the trial started and lasted until the maximum duration of 30 s was reached or until the dog looked away from the loudspeakers for more than 2 s. Subsequently the puppet's string was pulled again and once the dog oriented towards the loudspeakers the next trial started. During each trial the dog's looking behavior was scored in real time on the experimental laptop and the looking time (LT) was measured (i.e. total amount of time the dog spent looking in the direction of the loudspeakers).

Habituation phase. Dogs habituated on *feep*. During the habituation trials a *feep* token was repeated with an interstimulus interval of 1,300 ms. On each trial a different token was used (voices no. 1, 2, 3 and 4), which was assigned randomly to the trial. Dogs were exposed to a minimum of 6 and a maximum of 15 trials. The number of trials each participant was exposed to depended on their looking behavior. Once the total looking time on three consecutive trials reached 65% of the total looking time on the first three trials the habituation phase ended. It is assumed that the participant has then sufficiently been exposed to the stimulus as to lose interest. After the habituation phase the puppet was replaced by a new one in order to re-incite the dogs interest.

Test phase. During the test phase dogs were exposed to two types of test trials called *old* and *new* trials. On *old* trials dogs heard repeating alternations of an 'old' *feep* token (i.e. the token with voice no. 4 which was already used during habituation) and a 'new' *feep* token (token with voice no. 5). On *new* trials dogs were exposed to alternations of a completely new *faap* token (voice no. 5) and the 'new' *feep* token with voice no. 5, which was not used during habituation. The idea behind this is that if dogs (just as infants) show a significant increase in looking time on *new* trials, then this will be due to the change in vowel rather than due to the introduction of a new voice. On both types of trials the interstimulus interval was set to 1300 ms. A total of 14 trials were presented in the following pseudo-random order: *old, new, old, old, new, old, old, old, new, old, old, old, old, new* (as in Houston et al. 2007, see overview in Table 1).

Table 1.

Overview of Stimuli Presentation Sequence per Trial Type.

Trial type	Stimuli presentation sequence
Old	*feep* (V4), *feep* (V5), *feep* (V4), *feep* (V5) ...
New	*feep* (V5), *faap* (V5), *feep* (V5), *faap* (V5) …

4.1.5. Design

This study used a within-subjects design. The test trial type was the only independent, within-subject variable (*new* versus *old* trials). Looking time on each trial was measured as the dependent variable (defined as the total amount of time the dog spent looking in the direction of the loudspeakers during the trial).

4.2. Pilot experiment results and discussion

This method turned out to be inappropriate for dogs, and it became clear that it had to be adapted or even changed completely. No useful measurements could be made as I encountered the following problems.

Attention-getting. The puppet was not suitable for repeatedly drawing the dog' attention because after three or four trials, dogs lost interest in the puppet and would not orient towards the loudspeakers anymore. Considering that the minimum number of trials in the habituation phase was six this posed a problem. I decided that having a familiarization phase of fixed duration instead would be more appropriate (as is often done with infants and cotton-top tamarins, see for example Saffran et al. 2008 and Newport et al. 2004). In a familiarization phase participants are exposed to a stimulus for a fixed amount of time without the need for multiple trials (and drawings the dog's attention multiple times).

Interstimulus interval. The interstimulus interval of 1,300 ms seemed long as I observed dogs moving their eyes or pricking up their ears on the rhythm of the stimulus presentation. These pauses were so long that they allowed dogs to drop their ears or quickly glance away from the loudspeakers between stimulus presentation. When I reduced the interstimulus interval to 800 ms (as tested with the third participant) this reaction disappeared.

Number of test trials. Fourteen test trials proved far more than the dogs could handle: already after the three first trials dogs seemed to lose interest. I therefore concluded that the number of test trials had to be reduced drastically. I initially cut the amount down to two test trials (see section 5), then I added another two (four test trials in total), and finally I used seven test trials to see whether I could still obtain useful measurements on the last trials as well.

Stimulus presentation. Since I saw no signs of discrimination during the pilot tests I wondered whether something else was wrong. In a study on obedience to recorded commands done by Buytendijk and Fischel (1963, as cited in Miklósi 2007) dogs seemed to mostly rely on the beginning of a word in order to identify the right command. To exclude the possibility that dogs might only be listening to the beginning of a trial in order to decide whether it is interesting or not, I decided to opt for test trials where only one token is repeated instead of two tokens being alternated (ie. '*faap, faap, faap, faap...*' instead of '*feep, faap, feep, faap...*'; as in the extended variant of the Visual Habituation Procedure; Houston et al. 2007).

Scoring looking behavior. In this type of studies the scoring (i.e. logging the participant's looking behavior) is often done on line (i.e. in real-time during the test). In that case the experimenter is usually blind to the condition and listens to masking music during the experiment.

Since this method had never been used with dogs before for speech sound discrimination, I wanted to be able to hear the stimuli as well as see the dogs' reactions at the same time. This is of course less than ideal (experimenter bias), but it allowed me to make changes to the procedure for the next participant when necessary (for example reducing the interstimulus interval to a more appropriate level). Still, on line scoring was problematic as dogs sometimes responded to the sound of the buttons I pressed while scoring, and their behavior was not always immediately clear (i.e. had to be verified with the help of the recordings from the side-view camera). For all these reasons I decided to postpone scoring the videos until a few months after testing. This had the additional advantage of allowing me to use the video recordings from the side-view camera whenever the dog's behavior was unclear.

5. Experiment

As in the pilot study, the goal of this experiment was twofold: (1) to assess whether the chosen method and set-up is appropriate for dogs; (2) to assess whether dogs can discriminate between the non-words *feep* and *faap* (i.e. hear the difference between /e/ and /a/ which is a Dutch native contrast). The method used was based on the 'extended' variant of the Visual Habituation Procedure (Houston et al. 2007). However, the habituation phase was replaced by a familiarization phase of fixed duration (as often done in this type of research, see Saffran et al. 2008 and Newport & Aslin 2004). In the familiarization phase the dog listened to repetitions of *feep*. The test phase was slightly altered too. On *old* trials the dog was exposed to repetitions of the same non-word he/she had heard during familiarization (i.e. "*feep, feep, feep...*"). On *new* trials the dog was exposed to repetitions of the non-word *faap* instead (i.e. "*faap, faap, faap....*"). In both trial types the token used was the one not heard during familiarization (token with voice no. 5).

5.1. Method

5.1.1. Participants

Dogs were recruited in the same way as in the pilot. Seventeen dogs from Dutch-speaking homes and of various breeds were included in this study. Eight dogs were male and nine were female. Their average age was 4 years, 3 months old (ranging from 1 year, 1 month old to 9 years, 10 months old). An additional thirteen dogs were tested but excluded from analysis (drop out rate of 43% [5]) for the following reasons: continuous distraction by flies ($N = 10$), excessive barking or whining ($N = 2$), being scared during the experiment ($N = 1$). None of the dogs had earlier experience with similar experiments.

5.1.2. Stimuli

The stimuli were identical to those of the pilot study.

[5] If we exclude the dogs that were distracted by flies (which is clearly an environmental factor) from our calculation the drop-out rate is reduced to 15% (3/20).

5.1.3. Apparatus

The apparatus was identical to that of the pilot study.

5.1.4. Procedure

The procedure was identical to that of the pilot up until the start of the habituation phase. The habituation phase was replaced by a familiarization phase and the following changes were made:

Familiarization phase. During this phase dogs were exposed to repetitions of *feep* tokens (voices 1–4) in random order with an interstimulus interval of 800 ms during a fixed duration of time, which varied per participant: 3.5 min ($N = 1$) , 3 min ($N = 2$) and 2 min ($N = 14$) (see Table 2). The order of the *feep* tokens was randomized such that no token was repeated twice in a row. For the participants in the "puppet during familiarization" condition ($N = 3$) familiarization started once his/her attention was grabbed by the attention-getter (just as in the pilot). For those in the "no puppet during familiarization" condition ($N = 14$) familiarization started at an arbitrary moment determined by the experimenter when the dog was looking away from the loudspeakers (so as to see whether the sound would capture the dog's attention and make him/her orient towards the loudspeakers).

The test phase. As in the pilot, during the test phase dogs were exposed to two types of test trials called *old* and *new* trials. During this experiment however, on *old* trials dogs heard repetitions of a *feep* token that was not used during familiarization (token of voice no. 5; "*feep, feep, feep...*"), while on *new* trials dogs were exposed to repetitions of a *faap* token of the same voice (voice no. 5; "*faap, faap, faap...*"). On both types of trials the interstimulus interval was set to 800 ms. The number of trials as well as their order varied per participant (see Table 3).

In the "puppet during test phase" condition ($N = 9$) the trials started as soon as the dog, with the help of the attention getter (puppet), oriented towards the loudspeakers. In the "no puppet during test phase" condition ($N = 8$) the trial started at an arbitrary moment determined by the experimenter when the dog was looking away from the loudspeakers. This was done in order to see whether the sound would capture the dog's attention and make him/her orient towards the loudspeakers (which it did on 49 out of 64 trials). The trial duration varied per participant. In the "no fixed trial duration" condition ($N = 14$) the trials lasted until the maximum duration of 30 s was reached, or until the participant looked away from the loudspeakers for more than 2 s. In the "fixed trial duration" ($N = 3$) the trials had a fixed duration of 10 s (this was considered more appropriate given that the looking behavior was unclear at times or influenced by distractions in the environment – see Discussion in section 5.2).

Scoring videos. A few months later an observer who was naive to the details of the study and myself watched the muted video-recordings from the testing and registered the participants looking behavior. This was done by pressing one of two keys each time the dog changed his/her looking direction with software written by myself in ZEP (Veenker 2011). One key indicated that the dog was looking away from the loudspeakers, while the other key indicated that the dog was looking toward the loudspeakers. With this information the looking time (LT) during the familiarization phase as well as on each test trial was measured (i.e. the total time spent looking in the direction of the loudspeakers).

Table 2

Between-subject Variables

Familiarization duration	**No. of participants**
3.5 min	1
3 min	2
2.5 min	14
Use of attention-getter (puppet)	
At start of familiarization	3
At start of test trials	9
Never	5
Test trial duration	
Fixed (10 s)	3
Manual (2 s rule)	14

Table 3

Overview of Trial Order Among Participants.

No. test trials	**First trial type**	**Full test sequence**	**No. of participants**
2	*new*	*new–old*	3
2	*old*	*old–new*	1
4	*new*	*new–old–new–old*	11
7	*old*	*old–new–old–new–old–old–new*	2

5.1.5. Design

Again I used a within-subjects design where the test trial type (*old* vs. *new*) is the only independent variable. However, due to the exploratory nature of this study, I adjusted certain things while in the field which resulted in the following between-subject variables: familiarization duration, test trial duration, trial type tested first, number of test trials and use of attention-getter (see Table 2; for an overview of conditions per participant see Appendix C). Finally, just as in the pilot, looking time was the only dependent variable.

5.2. Experiment results and discussion

Varying looking time. The mean looking time on *old* trials (M = 1,435 ms, Mdn = 1,435 ms, SE = 233) was significantly less than the mean looking time on *new* trials (M = 4,687 ms , Mdn = 5,049 ms, SE = 611), $T = 151$, $p < 0.001$, $r = -.86$ (one tailed, Wilcoxon Signed Rank test). These differential looking times indicate that dogs are able to discriminate between the two words and as predicted, they spent more time looking during the *new* trials.

Interobserver reliability. The video recordings were scored by a naive observer and myself. Pearson's correlation coefficient was used to asses the correlation between the mean looking time scored on *new* trials and *old* trials by the second observer and myself. Scores for the average looking times (LT) on *new* (*faap*) trials strongly correlated: $N = 17$, $r = .64$, $p = .006$, as well as scores for the LT on *old* (*feep*) trials: $N = 17$, $r = .78$, $p < .001$ (according to Cohen 1988 when r is .50 the effect is large). Since the scores correlate, I used the average score from the two observers as input for all statistical analyses [6]. Also, this amount of correlation indicates that dogs looking behavior can fairly easily and consistently be interpreted.

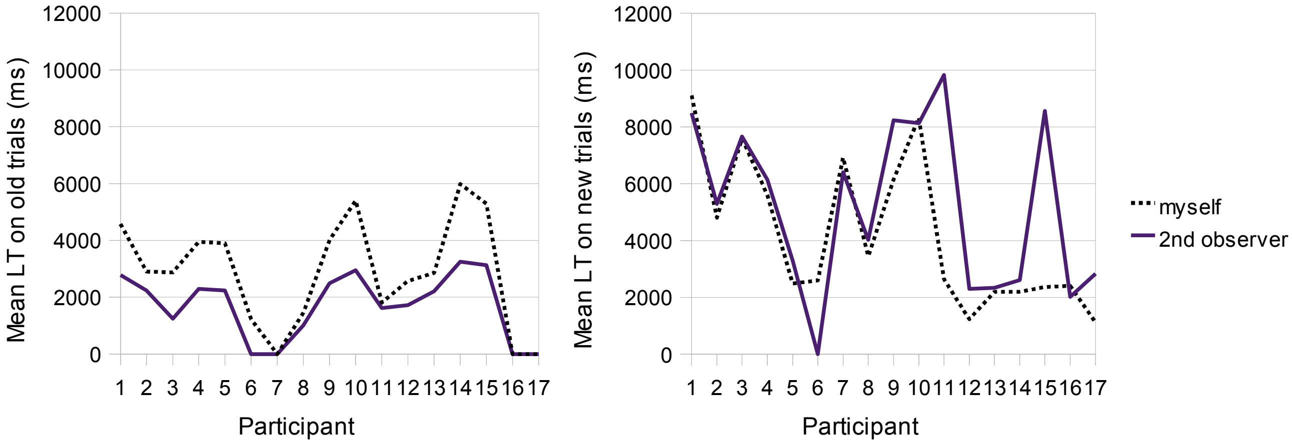

Figure 6. *Mean LT per trial type and participant as scored by the two observers.*

[6] Even when comparing the looking times as scored by each individual observer, the mean looking time on *old* trials remains significantly shorter than the mean looking on *new* trials ($T = 133$, $p < 0.001$, $r = -.84$ from the data collected by myself, and $T = 150$, $p < 0.001$, $r = -.84$ from the data collected by the 2nd observer; one-tailed Wilcoxon Signed Rank test)

Looking time – parametric data? As the sample size is very small, I found it hard to decide whether looking time is normally distributed in the population. If we look at the graphics of our samples (see Figures 7 and 8) they would seem to suggest that we are dealing with non-parametric data (especially mean LT on *new* trials). On the other hand, the Kolmogorov–Smirnov test tells us that the distribution of our sample does not significantly deviate from a normal distribution $D\ (17) = .101,\ p = 0.2$. Considering the small size of our sample however, this does not give us a definitive answer on the population distribution. Since the statistical power of a t-test is bigger than that of a non-parametric test I decided to use non-parametric tests (i.e. it is possible that a t-test detects a statistically significant difference while a non-parametric test cannot detect it). In order to judge this data in the strictest manner without making any assumptions on the population distribution, I have opted for using non-parametric tests.

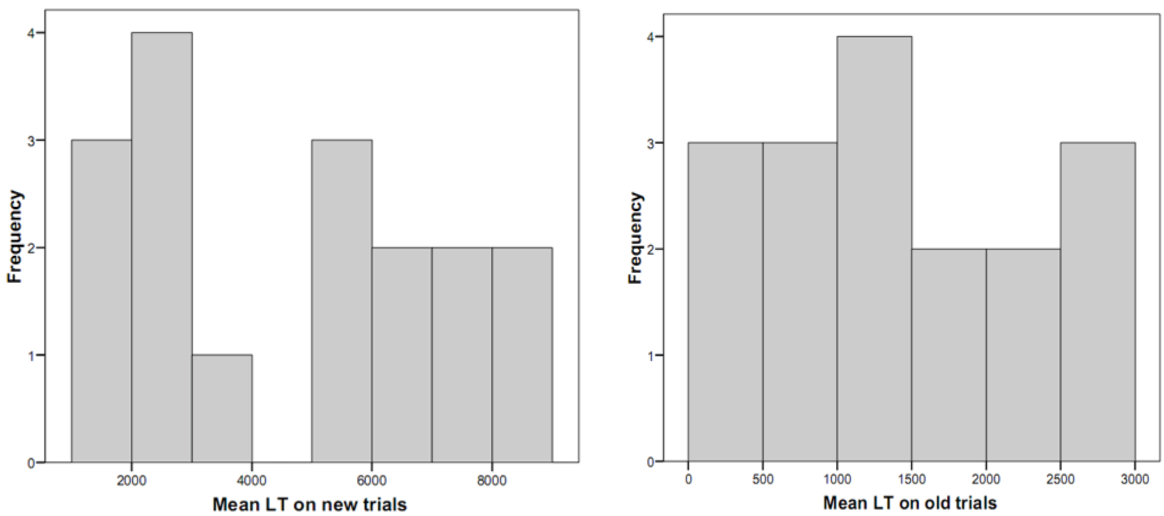

Figure 7. Histogram mean looking time (ms) on *new* and *old* trials.

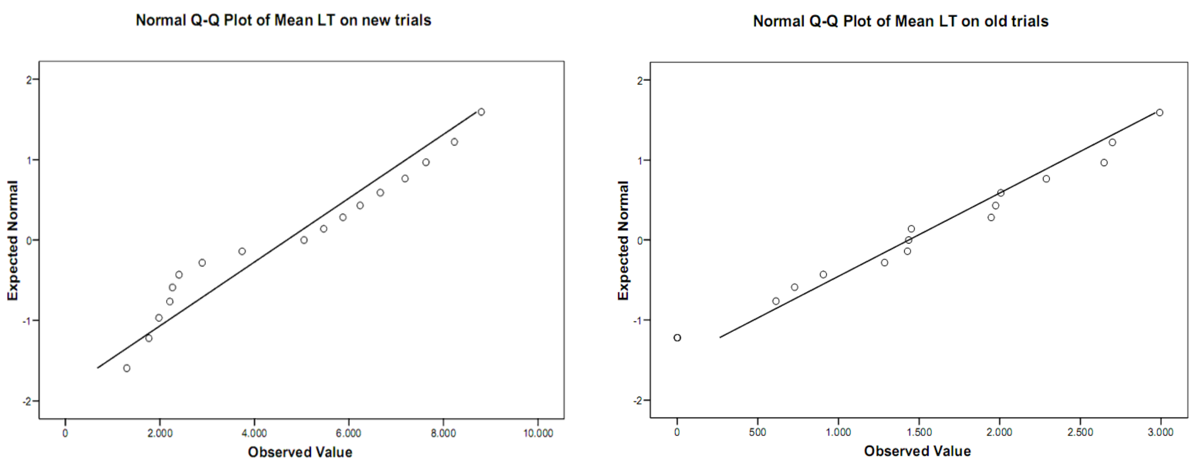

Figure 8. Q-Q plots for mean looking time (ms) on *new* trials and *old* trials.

The problem of counterbalancing. Unfortunately the planned testing on the 4th and 5th location had to be canceled. As a result I was not able to fully counterbalance the trial order across participants. Fifteen out of seventeen participants were presented with a *new* trials directly after the familiarization phase while only two participants were presented with an *old* trial directly after familiarization. One might thus argue that what we have in fact measured is the result of an increased interest during the first test trial due to the relatively long pause (silence) between the familiarization and test phase (M = 14 s), or the result of a boredom effect (looking time tends to decrease on subsequent trials).

Pause effect? To counter the first possibility I decided to exclude all first trials from analysis (i.e. the trials directly after the familiarization phase), as well as the participants who were only tested on two trials (4 participants excluded; 13 included). The results indicate that the mean looking time on *old* trials (M = 1275 ms, Mdn = 1286 ms, SE = 280) is still significantly less than the looking time on *new* trials (M = 4057 ms, Mdn = 2893 ms, SE = 670), T = 73, p = .03, r = -.53 (one tailed, Wilcoxon Signed Rank test). The effect this time is smaller but still large.

Boredom effect? Given that *new* trials always preceded *old* trials the above findings could be the result of a boredom effect. As the experiment progresses, the participant's interest decreases and looking time drops on each subsequent trial. To check whether the boredom effect could be the cause of the decreased looking time during *old* trials, I tested whether there was a significant decrease in looking time between the first and third trial (both *new* trials, N = 13) as well as between the second and fourth trial (both old trials, N = 11) trial. This does not seem to be the case (see Table 4), even though the difference is approaching significance. This might in fact be indicative of an overall decrease in looking time as the test phase progresses (see Figure 9). I also examined whether there is a correlation between trial number/order (i.e. 1st, 2nd, 3rd, 4th trial etc) and mean looking time. This does not seem to be the case either (rho = -.559, p = .096, Spearman, one-tailed).

Finally, if we look at the data of the two dogs that were tested on an *old* trial first we can see that these also follow the general trend with longer looking times on *new* trials (see Figure 10). In fact they both have zero looking time on *old* trials (both participants were in the "no puppet during test phase" condition).

Table 4.

Wilcoxon Test Results for Significant Decrease in Mean Looking Time as Trials Progress.

Trials compared	**n**	**T**	**p (one-tailed)**	**r**
1st and 3rd	13	45	.14	- .30
2nd and 4th	11	43	.06	- .48

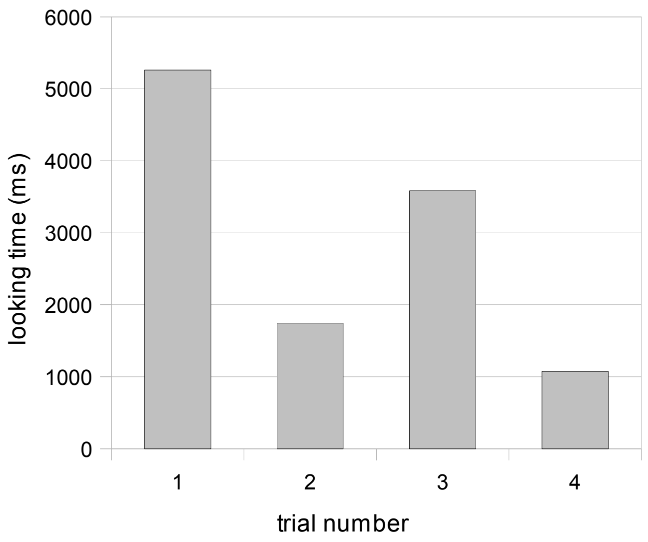

Figure 9. Mean looking time per trial. The 1st and 3rd trial are *new* trials. The 2nd and 4th trial are *old* trials.

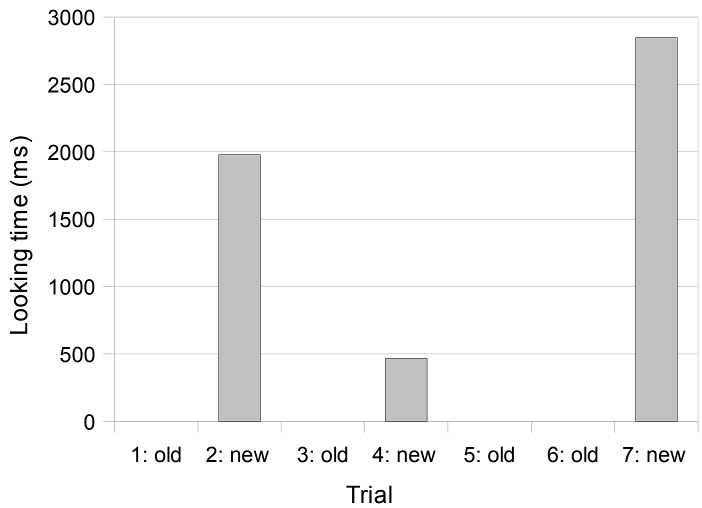

Figure 10. Mean looking time of the two dogs tested on an *old* trial first.

Puppet and looking time. Another question concerning the method is whether using the puppet (attention-getter) has an effect on the looking time. One would expect the mean looking time to be longer when the puppet is being used during the test phase ($N = 9$), than when it is not ($N = 8$). When not using the puppet, the dog has the choice to not look at the loudspeakers at all during a trial (which happened on 17 trials in total with 4 participants), while when using the puppet the dog is forced to look at the speakers even if only for an instant (a trial only starts upon the dog looking towards the puppet). The results show that the mean looking time on trials with a puppet ($M = 2966$ ms, $Mdn = 2698$ ms, $SE = 498$) did not significantly differ form mean looking time on trials without a puppet ($M = 3264$ ms, $Mdn = 3451$ ms, $SE = 653$), $U = 32$, $p = .74$, $r = -.09$ (Mann-Whitney Test).

Familiarization LT and test phase LT. Does the looking time during familiarization predict the mean looking time during test trials? Or does it perhaps predict the magnitude of the difference in mean LT between *old* and *new* trials (|mean LT on *new* trials – mean LT on *old* trials|) [7] ? No such correlations were found suggesting that looking time during familiarization cannot be used as a predictor for performance (see Table 5). This is of course assuming that a larger difference between mean LT on *old* and *new* trials indicates a better capacity for discrimination (which might of course not be true, see Discussion in section 6.2).

Table 5.

Pearson's Correlation Test for LT During Familiarization and (a) Mean LT on Trials (b) the Magnitude of the Difference Between Old and New Trials.

Familiarization time LT correlation with:	**N**	**r**	**p**
Mean LT per trial	17	.08	.75
Difference mean LT *old* and *new* trials*	17	.24	.37

* Calculated as: | mean LT new – mean LT old | (absolute value of subtraction)

[7] I used the absolute value of the difference between the mean LT on *new* trials and mean LT on *old* trials as a measure of individual performance (|mean LT on *new* trials – mean LT on *old* trials|). For more details and discussion on this see "measure of individual performance" in section 6.2)

No effect of gender. According to Müller et al. (2011) female dogs look longer at unexpected compared to expected events, whereas male dogs look equally long at both events. Considering a *new* trial as an unexpected event, one might expect that female dogs look longer on *new* trials than male dogs do. Or that the difference between looking time on *old* and *new* trials is larger for female dogs than for male dogs. None of the above expectations seem to be the case. (U = 29, p = .54 , r = -.16 and U = 31, p = .67 , r = -.12 respectively; Mann–Whitney two-tailed)

Questionnaire data. The online questionnaire was filled in for 12 out of the 17 participating dogs. Several measures were obtained from the questionnaire among which: (a) character traits such as playfulness, curiosity, intelligence, sociability, 'human-orientedness', (b) time spent with people, (c) time spent alone, (d) amount of time owner actively engages with dog, (e) amount of time owner actively engages with dog and talks to the dog, (d) lexicon size, and (e) quality of the relationship. No correlation between these measures and performance (i.e. difference in mean LT between *old* and *new* trials) was found (see Table 6). The age at which the dog was adopted into the family (puppy vs. adult) neither had an effect (U = 11, p = .73 , r = -.13), nor did the dog's training level (U = 13, p = .53 , r = -.21), or medication (four dogs were under medication for some ailment at the time of testing; U = 15, p = .93 , r = -.05; Mann–Whitney two-tailed). For details on how the questionnaire scores were calculated see Appendix B.

Table 6.

Correlation Between Several Questionnaire Scores and the Difference in Mean LT on Old and New Trials (Kendall's tau).

Questionnaire measure	**Difference in mean LT on old and new trials.***	
	tau_b	**p**
Playfulness	-.239	.327
Curiosity	.190	.436
Intelligence	.173	.487
Sociability	.264	.272
Human-orientedness	.047	.836
Time spent with people	.236	.297
Time spent alone	.066	.778
Time owner engages actively with dog	-.209	.375
Time owner engages actively with dogs **and** talks	-.240	.292
Lexicon size	-.260	.243
Quality of the relationship	.064	.780

* Calculated as: | mean LT new – mean LT old | (absolute value of subtraction)

6. Results & discussion

6.1. Results summary

The main aim of this study was to propose using the preferential looking paradigm for studying speech sound discrimination in dogs. To assess the validity of this method, I conducted a simple experiment in which I used this method to test whether dogs can discriminate between the Dutch native vowels /a/ and /e/ (something I expected to be within dogs capacities). Dogs were first familiarized with /e/ by listening to multiple tokens of the word *feep* and subsequently tested on trials where a *faap* token (containing the /a/) was presented, and on trials where a new *feep* token was presented. The results confirmed that dogs can discriminate between /a/ and /e/, and showed that the preferential looking paradigm can be used to study speech sound discrimination in dogs. This opens the way to also study discrimination of non-native Dutch vowels such as /æ/ and /ɛ/ (as in de Bree et al. 2010).

Furthermore, having conducted a pilot test before the experiment allowed me to fine-tune the method and make changes where appropriate (e.g. using a familiarization phase of fixed duration instead of habituation trials). Also, having different conditions such as the “puppet during test phase” and “no puppet during test phase”, allowed me to draw conclusions on the effect of each condition (e.g. the use of the attention getter had no influence on looking time). In section 6.3, I will discuss what I have learned in detail.

6.2. Limitations of the current study

Lack of a controlled environment. Since the tests were conducted with a mobile set-up at the dog daycare centers, the circumstances were less than ideal and some distracting factors could not be eliminated.

Noise. Invariably, on all locations, dogs could be heard barking in the distance and sometimes people could be heard walking outside the room. The latter problem was corrected by not allowing movements near the proximity of the testing room. The first problem, however, could not be solved. Depending on the dog’s personality, these noises caused either huge or moderate distraction which lead to exclusion from analysis or missing data on certain trials.

Smells. Dogs were often observed to intently smell around the area where the previous dog(s) had been seated. To reduce the amount of sniffing during testing, I let the dogs explore the room and sniff around for about 2 minutes before starting the experiment. Luckily, most dogs had lost interest in the smells by the end of the familiarization phase. Nonetheless, some dogs could be observed to be “picking up an interesting smell” all of a sudden, even though it had been in their proximity for a long time. This could perhaps have been eliminated if I had used cleansing materials.

Flies. Flies (fortunately only encountered on the third location) proved to be a serious distraction. Unfortunately, they could not be eliminated. This resulted in an extremely high drop-out rate for that location (79 %).

Lack of counterbalancing. Ideally, I would have counterbalanced the trial type order across participants and presented half of the participants with the *new* trial first and the other half with the *old* trial first. This however, did not happen as the planned experiments on the fourth and fifth location were canceled. In retrospect, I should have had the software randomly assign the participants to groups while the experiments progressed. Nevertheless, as I tried to argue in Section 5.2, the collected data remain useful, and indicate that dogs can indeed discriminate between *faap* and *feep.* Of course, this study should be repeated with proper counterbalancing.

Measure of individual performance. While we found evidence that dogs in general can discriminate *faap* and *feep*, the method applied does not provide us with the means to assess dogs’ individual ability (i.e. whether each dog looked significantly longer on *new* trials than on *old* trials). This would require having a bigger number of test trials (at least 12 or 14 like in Houston et al. 2007). Furthermore, we can ask ourselves whether there is any gradation in dog’s ability to discriminate between the two stimuli. Are there dogs that ‘perform better’, or is there no such thing as performance and are dogs simply either capable or incapable of discrimination? If performance could be measured in some way (given that performance is meaningful in this context), we could see whether it correlates to scores from the questionnaire. That is what I tried to do when I defined the absolute value of the difference between mean LT on *new* and *old* trials (|mean LT new trials – mean LT old trials|) as a measure for individual performance. The idea behind this is that, the larger the difference between LTs the better the dogs ability to discriminate (irrespective of the direction of that difference). However, this assumption is unfounded and therefore this measure might not accurately reflect performance. (It might for example reflect the level of surprise instead). Clearly, another way of measuring individual performance has to be found.

6.3. What have we learned? (or: how to test dogs)

The aim of this study was to determine whether the preferential looking paradigm, widely used with infants to study speech sound discrimination, can be used for studying the same capacity in dogs. The brief answer is: yes. Even though details of the method need to be adjusted to better suit dogs, this paradigm can certainly be used. Below I highlight some of these (potential) adjustments.

Habituation vs. familiarization. As explained in Section 4.2, habituation proved inappropriate for reasons either inherent to dogs personalities, or to my set-up and the noise in the environment. It was therefore replaced by a familiarization phase. It is my personal opinion that it matters little which of the two methods is use. The essential idea behind both is the same: the participants are exposed to a stimulus until they 'get used to it' and loose interest. The difference is that with habituation trials the researcher can measure the 'drop in attention' by the drop in looking time on subsequent habituation trials. When a certain degree of drop in attention is reached (called the habituation criterion), it is assumed that the participant has been sufficiently exposed to the stimulus and habituation ends. When using familiarization, the researcher chooses a fixed length of time during which the participant is exposed to the stimulus. In the first case the researcher hopes that the habituation criterion chosen is meaningful and that it does indeed reflect that the participant has 'gotten used to' the stimulus. In the later, the researcher hopes that the familiarization duration chosen is sufficient to make the participant 'get used to' the stimulus. Given that the puppet's capacity to attract attention repeatedly is limited I suggest using familiarization instead of habituation (or else finding a more sophisticated and engaging attention-getter).

Fixed trial duration. The 'looking away for 2 seconds' criterion for signaling a decrease in attention is in my opinion inappropriate for dogs or for noisy environments (such as the one I tested in). The dogs often seemed to get distracted and look away from the speakers only to recapture their interest and stare intently at the speakers a few seconds later. This might have been due to some environmental factor (i.e. noises or lingering smells might have distracted the dogs). Or perhaps it has something to do with these dogs' personalities (guess: impatience, quickly bored, need for human attention/comfort). I often observed dogs nervously/impatiently "fidgeting around", turning their heads away from the speakers, but without in fact losing interest in the sound (as their subsequent glances and stares towards the speakers suggest). I therefore think that giving the test trials a fixed duration is more appropriate.

Utility of the attention getter. As mentioned before, I also questioned the utility of the puppet (attention-getter) and therefore did half of the tests without the puppet. The later worked equally well as the dogs naturally oriented towards the speakers upon hearing the stimuli. The only drawback in this case was that the experimenter had to wait until the dog looked away from the loudspeakers before starting the new trial (so as to see whether the dog would orient towards them). One could argue that this method leads to a less controlled procedure as the time between trials varied. This problem could perhaps be alleviated by using an additional attention-getter that is located away from the speakers. At the same time this could lead to additional distraction (or once again, a more sophisticated attention-getter should be found). From what I observed I draw the conclusion that the puppet functions as an additional element of distraction, and that if possible should be eliminated. In fact, two of the dogs tested got scared by the puppet's sudden movement. Statistically speaking however, the attention-getter had no influence on the looking time.

Number of test trials and familiarization duration. While 14 test trials are clearly too much for dogs I still managed to get meaningful measurements for a maximum of 7 trials (two participants). The maximum and optimal number of test trials remains to be explored. Furthermore, a familiarization phase of 2.5 min proved sufficient. Given that the average amount of time dogs spent looking at the loudspeakers during familiarization is 36 s (ranging from 11 s to 85 s), the duration of the familiarization phase could possibly be lowered.

Stimulus presentation on test trials. In the pilot, two tokens alternated during test trials while in the experiment a single token was repeated during test trials. Whether we would have obtained the same experimental results with alternating trials remains an open question. As far as infants are concerned, the differences in looking time are even sharper when alternating trials are used (Houston et al. 2007). It is unknown whether this is true for dogs as well. The failure of the pilot study might not have been due to the trial type but to other factors (e.g. habituation, interstimulus interval).

Interstimulus interval. When the interstimulus interval was reduced from 1,300 ms to 800 ms dogs stopped pricking up their ears on the rhythm of the stimulus presentation. While I initially considered this to be positive it is unknown whether a different interstimulus interval would have affected the results. Could the pricking up of ears be a sign of sustained attention or surprise? Would a longer interstimulus interval have led to better results, or would it just have confused the dogs?

6.4. Other observations

I would like to hereby share some of my other observations made during testing. Beside the video recordings themselves I do not always have measurements to back these up, so please take them with a grain of salt. Nevertheless I think there is some value in sharing these observations as they could lead to ideas for further research. Also allow me, at this early experimental state, to at times insert my own speculative interpretations of these observed behaviors.

Looking time: a good measure of auditive attention? Dogs seem to naturally look in the direction of the sound source when intrigued by a sound. Indeed, out of the 33 dogs tested only 1 seemed to have no interest whatsoever for the sound (beside also being distracted by flies) and did therefore not display any head movements towards the loudspeaker. Looking time would thus seem to be a good measure, even though the duration itself might vary greatly among dogs (ranging from 11 s to 85 s during familiarization).

Filtering out uninteresting noises. Some of the behaviors I observed gave me the impression that dogs might be very good at filtering out “uninteresting” noises. None of the dogs reacted to a *beep* sound that I used in the hope to re-gain their attention when moving the puppet solicited no reaction (in the pilot study). In fact they did not even seem to be hearing it. Could this also be the case with other electronic noises? Similarly, there seemed to be a curious delay (M = 2,843 ms) between the trial onset and the dogs’ reaction. At this magnitude this delay has nothing to do with the dogs natural reaction time. I suspect it is more a matter of interest or attention. To me it seems as if the dogs had decided that nothing interesting was going to come out of the loudspeakers and had therefore concentrated on something else (like listening to a distant sound, intently investigating a smell or expectantly looking at the door), or simply started to doze off. But as the stimulus presentation continued, it seemed as if the dogs had to “change their mind” and take a moment to pay attention again. This reminds me of the cocktail party effect in humans where the listener focuses on what is interesting for him/her without however totally excluding other noises from auditory processing.

Sensitive to intonation and voices. Some of the *feep* tokens seemed to be more salient than others as some of the dogs (about three) would prick up their ears and/or make eye movements every time these were presented. As far as I can remember these were the tokens that were also more salient to me (i.e.: one had the lowest voice and the other had a more “lively” intonation). It is commonly known that dogs are very sensitive to intonation and perhaps this would reflect in their looking times as well.

Sensitive to sequential properties (?) Due to inadequate functioning of the randomization method in the software during familiarization a particular *feep* token was sometimes repeated twice in a row (this problem was solved in later trials). A couple of dogs seemed to be very sensitive to this and could be seen pricking up their ears every time they heard the repetition of the same token. This would mean that after such a brief exposure they could easily tell that the two tokens were identical. Perhaps this means that dogs also pay attention to sequential properties. It would therefore be very interesting to see how dogs perform in rule learning (e.g. their ability to discriminate between ABB and AAB types of sequences).

6.5. Future work

This study can serve as a pilot to illustrate that there is a simple way of testing speech sound discrimination in dogs, which does not require training. Furthermore, since dogs grow up and live in linguistically rich environments, they are ideal for comparison with infants. Listening to human language is not something that dogs need to get used to in a lab. Dogs are exposed to language on a daily basis and are even being spoken to. Given that: (a) dogs are easy to recruit (they are where humans are after all); (b) this type of research requires minimal material; and (c) I personally perceived a lot of enthusiasm in the dog owner community, I heartily encourage developing this research further.

This study has managed to yield some interesting results, nevertheless, I highly advise repeating the same experiment under more controlled conditions: by using a sound attenuated room/booth, properly counterbalancing participants, minimizing the between-subject variables etc.. Once that is done, one can decide to either focus on this method and try to optimize it or immediately go on and attempt to answer another question around speech sound perception in dogs.

6.5.1. Optimize the preferential looking paradigm for dogs

For optimizing this method, I advise testing the effects of familiarization duration, trial duration, interstimulus interval and use of the attention-getter, as well as defining the optimal number of test trials. While doing this, it would be a good idea to use stimuli that dogs can readily discriminate. The non-words, *faap* and *feep,* used in this study, differ only in one vowel. Although the evidence shows that dogs can distinguish them, it is not impossible that it might have been a challenge for some of the dogs. By using non-words that differ even more, or perhaps also vary in intonation (which dogs seem to be very sensitive to), one can make sure that what is being tested is the set-up and not the dogs' ability to discriminate.

6.5.2. Follow-up study

In order to parallel the research that is currently being conducted at the Babylab in Utrecht (de Bree et al. 2010), one could follow up this study by testing Dutch dogs' sensitivity to a non-native contrast, using the same materials (i.e. test the non-native /æ/ and /ɛ/ contrast). Infants lose the ability to hear the difference between non-native contrasts during their first year of life. Will dogs also show a similar 'native language' effect? If so, at what age does this appear, and how does this develop? Lalonde and Werker (1995) point out that the ability of human infants to discriminate non-native vowels declines earlier (at 6 months old) than the ability to discriminate non-native consonants (at 10–12 months old). Would puppies show a similar pattern?

6.5.3. Other research questions

Of course we do not need to limit ourselves to phoneme discrimination. As I already alluded to in the introduction, other species are widely used in linguistic research that centers around categorical perception and statistical learning. We can draw our inspiration from those areas, as well as from psycholinguistics with infants. Below I highlight the different directions "canine linguistic research" could move towards.

Preference for human speech. One of the most basic questions to ask is whether dogs and puppies show a preference for human speech over other sounds just like human infants do (Colombo & Bundy, 1981; Glenn et al., 1981). Furthermore, people often speak 'baby talk' to their dogs (which is termed 'doggerel'). Infants prefer listening to 'baby talk' (termed motherese) over adult directed speech (Fernald & Kuhl, 1987), and it is hypothesized that this is due to "innate predispositions" and the association with pleasant care-taking situations. Given the striking similarity between motherese and doggerel (Hirsh–Pasek and Treiman, 1982), we can ask ourselves whether dogs show a similar preference for doggerel. If so, why and what would that mean for our hypothesized assumptions about the infant's preference?

Language discrimination. The ability to discriminate between languages also raises questions. Human newborns, cotton-top tamarins (a monkey) and rats have been shown to discriminate between Dutch and Japanese, two languages belonging to different rhythmic categories (Ramus et al. 2000; Toro et al. 2005). However, when two languages belong to the same rhythmic category infants cannot discriminate them unless one is their native language (Nazzi et al. 1998; Mehler et al. 1988). What about dogs? I expect dogs to be able to discriminate between languages of different rhythmic categories, just as rats and cotton-top tamarins do. But what about languages from the same rhythmic category? Will dogs, just like infants, be more capable of distinguishing their 'native' language?

Speech segmentation. Unlike popular belief, speech has no pauses between words but is a continuous stream of phonemes. It therefore has to be segmented in order to be understood (which is something our brain does for us wonderfully, we call this speech segmentation). Rats have been shown to segment speech as well but seem to rely on a slightly different strategy (i.e. relying more on the frequency of syllable co-occurrence than transitional probabilities, Toro & Trobalon 2005).

Cotton-top tamarins can also segment speech but it still unclear how they do this (Hauser et al. 2001). Given that the Border Collie Chaser could discriminate between the words for objects and commands and combine these, it seems very probable that dogs also segment speech. The question is how.

Rule learning. Turning to the somewhat more complex question of aptitude for grammar, we find that rats cannot distinguish between ABB and AAB patterns (i.e. after a syllable of type A, a syllable of type B follows twice, versus: after two times syllable A, syllable B follows), neither can they learn non adjacent dependencies of the type AxB (i.e. that a syllable of type A is always followed by a syllable of type B with the interval of a random syllable 'x') (Toro & Trobalon 2005). Cotton-top tamarins however can learn $(AB)^n$ but not A^nB^n patterns (Fitch & Hauser 2004) and seem to be able to learn some type of non-adjacent dependencies (Newport & Aslin 2004). The question here would be: how do dogs fare?

Predictive dependencies in language. Researchers claim that language has a structure that is attuned to human learning abilities. Human languages have a predictive phrase structure which means that the presence of a certain word type predicts the presence of another word type in the same phrase. To use the example of Saffran et al. (2008): when an article (*a* or *the*) occurs in a phrase then it is very likely to find a noun somewhere down-stream in the phrase. However, the presence of a noun does not require the presence of an article. We call these type of word dependencies predictive. Saffran et al. (2008) showed that when infants as well as cotton-top tamarins are exposed to an artificial language containing such predictive dependencies (a predictive language), they can easily pick out the ungrammatical sentences when tested. However, when the language does not contain such dependencies (the language is non-predictive), they fail to do so. These predictive structures are not only easier for humans to learn but for cotton-top tamarins too. It would therefore be interesting to repeat this experiment with dogs and see whether these structures are easier to learn for a non-primate as well. There was one difference, however, in the abilities of infants and cotton-top tamarins. When these predictive dependencies had to be generalized over word classes (which requires learning the dependencies between word categories instead of individual words) cotton-top tamarins failed to discriminate grammatical from ungrammatical sentences. Cotton-top tamarins are apparently unable to create word categories, or they lack the memory to do so. Given that Chaser (the famous Border Collie) could understand object categories and link them to words, it would be very interesting to see how she would fare on such a test.

Detecting ungrammaticality. Finally, we can also exploit the fact that dogs grow up and live in linguistically rich environments by testing whether they can identify ungrammatical phrases in everyday language. Such a study was done with infants by Santelmann & Jusczyk (1998), who tested whether infants are aware of the "*is* ….. *-ing*" dependency in English. The results showed that 18-month old infants have a preference for well-formed over ill-formed phrases containing this dependency, while 15-months-olds do not. Since this has a lot to do with "keeping count" (or statistical learning), and dogs are exposed to language on a daily basis, we could test them on a similar tasks.

7. Conclusion

This study aimed to answer two questions: (1) whether the preferential looking paradigm, a paradigm widely used in infant studies, can be used to study speech sound discrimination in dogs; and (2) whether dogs that have grown up in Dutch-speaking homes can discriminate between the Dutch vowels /a/ and /e/ (embedded in the non-words *faap* and *feep*).

The results show that the preferential looking paradigm can indeed be used to study speech sound discrimination in dogs. Beside allowing for a direct comparison with infants, this method has the additional benefits of being non-invasive and quick. Furthermore, it requires only minimal material and no training.

The results also show that dogs raised and living in Dutch speaking homes can distinguish between the Dutch vowels /a/ and /e/ (embedded in the non-words *faap* and *feep*). I used the preferential looking paradigm to test this. Seventeen dogs were first familiarized with /e/ by listening to multiple tokens of the word *feep.* Subsequently, the dogs were exposed to two types of trials: (a) *new* trials during which a *faap* token was repeated, and (b) *old* trials during which a *feep* token was repeated. During both types of trials, the dogs' looking times were measured and compared. Dogs looked significantly longer during *new* trials than during *old* trials, indicating that they do perceive a difference between /a/ and /e/. Given that dogs can discriminate a Dutch native contrast we can use the same method to test whether dogs can discriminate non-native vowels such as /æ/ and /ɛ/ (paralleling the infant study done by de Bree et al. 2010). If dogs from Dutch-speaking homes cannot discriminate these 'non-native' vowels (just like 6 and 8 month old Dutch infants cannot), then dogs will be the first species shown to readily and naturally (i.e. without prolonged exposure in a lab to a limited set of stimuli) categorize speech sounds from the ambient language.

Of course we do not need to limit ourselves to phoneme discrimination. As I highlighted in section 6.5.3., we can take our inspiration from the numerous studies being done with human infants, non-human primates and rodents. From testing preference for 'doggerel' to the learning of 'grammatical rules', a world of opportunities is awaiting.

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.

Baru, A.V. & Shmigidina, G.N. (1977). *Role of the auditory cortex in recognition of synthesized vowels by dogs.* Neuroscience and Behavioral Physiology, 8, 3, 197–204.

Belyaev, D.K. (1978). *Destabilizing selection as a factor in domestication.* Journal of Heredity, 70, 301–308.

Bloom, P., (2004). *Can a dog learn a word?* Science 304, 1605–1606.

de Bree, E., Kerkhoff, A., de Klerk, M., Gerrits, E., Wijnen, F. (2010). *Speech sound discrimination using hybrid visual habituation with multiple speakers.* Poster presented at the International Conference on Infant Studies, Baltimore, MD.

Burdick, C.K. & Miller, J.D. (1975). *Speech perception by the chinchilla: discrimination of sustained /a/ and /i/.* The Journal of Acoustical Society of America, 58, 2, 415–27.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic Press.

Colombo, J. & Bundy, R.S. (1981). *A method for the measurement of infant auditory selectivity.* Infant Behavior and Development, 4, 219–223.

Davis, S.J.M., & Valla, F.R. (1987). *Evidence for domestication of the dog 12,000 years ago in the Natufian of Israel.* Nature 276, 608–610.

Faragó, T., Pongracz, P., Miklósi, Á., Huber, L., Viranyi, Z., & Range, F. (2010). *Dogs' expectation about signalers' body size by virtue of their growls.* PLoS ONE 5,12.

Fernald, A. & Kuhl, P. (1987). *Acoustic determinants of infant preference for motherese speech.* Infant Behavior and Development, 10, 279–283.

Fitch,W.T. & Hauser, M.D., (2004). *Computational constraints on syntactic processing in a nonhuman primate.* Science, 303, 5656, 377–380.

Gentner, T.Q., Fenn, K.M., Margoliash, D., Nusbaum, H.C. (2006). *Recursive syntactic pattern learning by songbirds.* Nature, 440, 1204–1207.

Glenn, S.M., Cunningham, C.C., Joyce, P.F. (1981). *A study of auditory preferences in non-handicapped infants and infants with Down's Syndrome.* Child Development, 52, 1303–1307.

Hare, B. & Tomasello, M. (2005). *Human-like social skills in dogs?* Trends in Cognitive Sciences, 9, 9, 439–444.

Harnad, S. (1987). *Psychophysical and cognitive aspects of categorical perception: A critical* *overview.* Harnad, S. (Ed.) *Categorical perception: the groundwork of cognition.* New York: Cambridge University Press.

Hauser, M.D. (1996). *The evolution of communication.* Cambridge: Bradford Books/MIT Press.

Hauser, M.D., Newport, E.L., Aslin, R.N. (2001). *Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins.* Cognition, 78, 3, B53–B64.

Hirsh-Pasek, K. & Treiman, R. (1982). *Doggerel: motherese in a new context.* Journal of Child Language, 9, 229–237.

Houston, D.M., Horn, D.L., Qi, R., Ting, J.Y., Gao, S. (2007). *Assessing speech discrimination in individual infants.* Infancy, 12(2), 119–145.

Howell, T.J., Conduit, R., Toukhsati, S., Bennett, P. (2011). *Auditory stimulus discrimination recorded in dogs, as indicated by mismatch negativity (MMN).* Behavioural Processes, 89, 1, 8–13.

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

Kuhl, P.K. (2004). *Early language acquisition: cracking the speech code.* Nature Reviews Neuroscience 5, 831–843.

Kuhl, P.K. & Miller, J.D. (1975). *Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants.* Science 190, 4209, 69–72.

Kuhl, P.K. & Padden, D.M. (1983). *Enhanced discriminability at the phonetic boundaries for the place feature in macaques.* The Journal of the Acoustical Society of America, 73, 1003–1010.

Kubinyi, E., Virányi, Zs., Miklósi, A. (2007). *Comparative social cognition: from wolf and dog to humans.* Comparative Cognition & Behavior Reviews, 2, 26–46.

Lalonde, C.E. & Werker, F.W. (1995). *Cognitive influences on cross-language speech perception in Infancy.* Infant Behavior and Development 18, 459–475.

Miklósi, Á. (2007). "Dog behavior, evolution, and cognition". New York: Oxford University Press.

Miklósi, Á., Kubinyi, E., Topál, J., Gácsi, M., Virányi, Zs., Csányi, V. (2003). *A simple reason for a big difference: wolves do not look back at humans, but dogs do.* Current Biology, 13, 9, 763–766.

Miklósi, Á., Topál, J., Csányi, V. (2004). *Comparative social cognition: what can dogs teach us?* Animal Behaviour, 67, 6, 995–1004.

Mehler, J., Jusczyk, E.W., Lambertz, G., Halsted, N., Bertoncini, J., Amiel–Tison, C. (1988). *A precursor of language acquisition in young infants.* Cognition, 29, 143–178.

Müller, C.A., Mayer, C., Dörrenberg, S., Huber, L., Range, F. (2011). *Female but not male dogs respond to a size constancy violation.* Biology Letters, 7, 5, 689–691.

Newport, E.L. & Aslin, R.N. (2004). *Learning at a distance I. Statistical learning of non-adjacent dependencies.* Cognitive Psychology 48, 2, 127–162.

Newport, E. L., Hauser, M.D., Spaepen, G., Aslin, R.N. (2004). *Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate.* Cognitive Psychology, 49, 2, 85–117.

Pilley, J.W. & Reid, A.K. (2011). Border collie comprehends object names as verbal referents. Behavioural Processes, 86, 2, 184–195.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., Mehler,. J. (2000). *Language discrimination by human newborns and by cotton-top tamarin monkeys.* Science, 288, 5464, 349–351.

Saffran, J. R. (2002). *Constraints on statistical language learning.* Journal of Memory and Language, 47, 1, 172–196.

Saffran, J.R., Hauser, M.D., Seibel, R., Kapfhamer, J., Tsao, F., Cushman, F. (2008). *Grammatical pattern learning by human infants and cotton-top tamarin monkeys.* Cognition, 107, 2, 479–500.

Saffran, J. R., Pollak, S.D., Seibel, R.L., Shkolnik, A. (2007). *Dog is a dog is a dog: infant rule learning is not specific to language.* Cognition, 105, 3, 669–680.

Santelmann, L.M. & Jusczyk, P.W. (1998). *Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space.* Cognition, 69, 2, 105–134.

Svartberg, K. & Forkman, B. (2002). *Personality traits in the domestic dog (Canis familiaris).* Applied Animal Behaviour Science 79, 133–155.

Svartberg, K., Tapper, I., Temrin, H., Radesäter, T., Thorman, S. (2005). *Consistency of personality traits in dogs.* Animal Behaviour, 69, 2, 283–291.

Topál, J., Miklósi, Á., Gácsi, M., Dóka, A., Pongrácz, P., Kubinyi, E., Virányi Zs., Csányi, V. (2009). *The dog as a model for understanding human social behavior.* Advances in the Study of Animal Behaviour, 39, 71–116.

Toro, J. M. & Trobalon, J.B., (2005). *Statistical computations over a speech stream in a rodent.* Perception & Psychophysics, 67, 5, 867–875.

Toro, J. M., Trobalon, J.B., Sebastian–Galles, N. (2005). *Effects of backward speech and speaker variability in language discrimination by rats.* Journal of experimental psychology. Animal Behavior Processes, 31, 1, 95–100.

Vas, J., Topál, J., Gácsi, M., Miklósi, Á., Csányi, V. (2005). *A friend or an enemy? Dogs' reaction to an unfamiliar person showing behavioural cues of threat and friendliness at different times.* Applied Animal Behaviour Science, 94, 99–115.

Veenker, T. J. G. (2011). *The Zep Experiment Control Application (Version 0.6.12)* [Computer software]. Utrecht Institute of Linguistics OTS, Utrecht University. Retrieved *20-04-2011*. Available from http://www.hum.uu.nl/uilots/lab/zep/

Vila, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E., Honeycutt, R.L., Crandall, K. A., Lundeberg, J., Wayne, R.K. (1997). *Multiple and ancient origins of the domestic dog.* Science 276, 5319, 1687–1689.

Virányi, Zs, Gácsi, M., Kubinyi, E., Topál, J., Belényi, B., Ujfalussy, D., Miklósi, Á. (2008). *Comprehension of human pointing gestures in young human-reared wolves (Canis lupus) and dogs (Canis familiaris).* Animal Cognition, 11, 3, 373–387.

West, R.E. & Young, R.J. (2002). *Do domestic dogs show any evidence of being able to count? Animal Cognition* 5, 3, 183–186.

Wijnen, F. (in press). *Acquisition of linguistic categories: cross-domain convergences.* In: J. Bolhuis & M. Everaert (Eds.), Birdsong, Speech and Language: Converging Mechanisms. Cambridge MA: MIT Press.

Appendix A

Questionnaire

1. About your dog

Goal: *get general information about the dog.*

1 Name […]
2 Breed [...]
3 Gender & fertility[m/m neut/f/f neut]
4 If female and not neutered:
  4.1 Every how many months does your dog come in season? [number]
  4.2 When was the last time your dog was in heat? [date]
5 Birthday [date]
6 How old was your dog when you acquired him/her? (ex. 8 weeks or 3 months or 2 years)...
7 Where does your dog come from? [list]
  7.1 bred myself
  7.2 from breeder
  7.3 from animal shelter
  7.4 petshop
  7.5 previous owner
  7.6 found (on street)
  7.7 other
8 Why did you acquire your dog?
  8.1 Company
  8.2 Breeding/showing
  8.3 Sports (field trials/ hunting etc)
  8.4 Work (guide dog, therapy dog etc)
  8.5 Other [...]
9 Has your dog followed any courses? (ex: puppy courses, agility training etc) [y/n]
  9.1 If so, what kind?

2. Health & behaviour

Goal: *See if there are any medical reasons to exclude this dog from the study.*

1 Does your dog have sight problems? [y/n]
  - If yes, please specify: [...]
2 Does your dog have hearing problems? [y/n]
  - If yes, please specify: [...]
3 Does your dog currently have any other health problems? [y/n]
  - If yes, which? […]
4 Does your dog currenlty use any medicines? [y/n]
  - If yes, which? […]
5 Does your dog have any serious behavioural problems? (such as fear, aggression, excessive barking etc) [y/n]
  - If yes, please specify: [...]

3. Character
Goal: Get a general impression of the dog's character (or the owner's judgment of it).
(Scale 0–5: not at all – a lot)

Playfull	[scale]
Curious	[scale]
Chase-prone	[scale]
Social	[scale]
Intelligent	[scale]
Tendency to bark/make noises	[scale]
Active / energetic/ always on the go	[scale]
Aggressive	[scale]
Shy	[scale]
Motivated by food	[scale]
Motivated by toys	[scale]

***Note:** I used the personality traits developed by Svartberg&Forkman 2002 and Svartberg et al. 2005 and adapted some questions from the C-BARQ (Canine Behavioral Assessment and Research Questionnair)*

4. Human-oriented
***Goal:** Get a rough indication of how "human oriented" the dog is.*
(Scale 0–4: never-always)

1 My dog seems to attend/listen closely to everything I say or do.
2 My dog tends to follow me (or other members of the household) about the house, from room to room.)
3 My dog tends to sit close to, or in contact with, me (or others) when I am sitting down.
4 My dog actively seeks attention (blaffen, met spelletje naar uw toe komen enz)
5 My dog becomes agitated (whines, jumps up, tries to intervene) when I (or others) show affection for someone else.
6 My dog likes meeting new people and has a lot of interest in them.
7 My dog enjoys being petted by me (or a family member).
8 My dog enjoys being petted by strangers.
9 My dog joyfully accepts to play when I (or a family member) invite him/her to play. [scale]
10 My dog joyfully accepts to play when a stranges invites him/her to play. [scale]

***Note:** Used and adapted questions from the C-BARQ.*

5. Time spent

1 On a typical day, how many hours per day is your dog:
1.1 Actively involved with people? (ie. playing, being trained or walked etc)
1.2 Near people?(in same room and space where the dog can see and hear people, without being actively involed with them) [number]
1.3 With other dogs or animals without people around. [number]
1.4 Alone. [number]

6. Language exposure
***Goal:** language exposure of the dog.*
(Scales used 0–4 and 0–7: never–always)

1 Which language is mainly being spoken in the dog's presence? [dutch/dialect + .../other +…]
2 Is there an other language that your dog is regularly exposed to? *We are not concerned with a*

few sentences or commands here but whole conversations in an other language that your dog is regularly witnessing. [y/n]

2.1 If yes:
  2.1.1 Which language? [….]
  2.1.2 How many hours per week? [number]
  2.1.3 From whom does your dog hear this language? [choose]
    2.1.3.1 From someone that your dog has a close relationship with (i.e. family member, dog sitter etc).
    2.1.3.2 From someone your dog knows.
    2.1.3.3 From strangers
    2.1.3.4 Other [...]
  2.1.4 In which context does your dog hear this language? [choose]

3 The dog is being talked to in this language.
4 The dog hears people in the same space speaking this language.
5 The dog hears this language in his surroundings (ex: in park, in shops, from neighbors etc)
6 Other: [...]
7 How many hours per week are you personally actively engaged with your dog? (ie. playing, walking, talking, training) [number]
8 When you are actively engaged with your dog, how often do you talk to him/her? [scale 0–7]
9 When you talk to your dog:
  9.1 How often do you talk in commands? [scale 0–4]
  9.2 How often do you use simplified speech with a “cooing” pattern of intonation as if you were talking to an infant or toddler? [scale 0–4]
  9.3 How often do you use normal speech, as if adressing an adult human? [scale 0–4]

7. Lexicon

Goal: *Establish the dog's “lexicon”*

(Scale 0–5: never–always)

1 Which commands can your dog respond to? (ex: sit, stay, get etc) […]
  1.1 How many commands are these? [number]
2 Which other words does your dog know? (i.e. names of objects and people: ball, stick, car, food, home, daddy etc) […]
  2.1 How many words are these?
3 Beside the commands, are there any other sentences your dog responds to? (ex: “let's go walking” , “where is the ball?”, “what have you done?”): […]
  3.1 How many sentences are these? [number]
4 When your dog react to what you say, you have the impression that he/she:
  4.1 reacts to **what** I say (the actual words) [scale 0–5]
  4.2 reacts to the **way** I say it (intonation, gestures etc) [scale 0–5]

8. Communication

Goal: *get an indication of the quality of the dog-owner relationship.*

(Scale 0–4: not at all – very much)

6 How do you judge your relationship with your dog? [scale]
7 How obedient is your dog? [scale]
8 How often does your dog display behaviour that you find annoying or unpleasant? (i.e. whinning excessively, being too active or agressive etc)
9 How well does your dog understand your mood? [scale]
  - Do you have examples of this?

10 Does your dog ever try to communicate or grab your attention? (i.e. coming towards you to

be petted, bark when he wants to go out). [scale]

• If so, what does your dog do? […]

Note: *The responses will probably be extremely subjective but I think they nevertheless serve their purpose in that they tell us something about how the owner perceives (and possibly treats) his dog.*

9. About you

Goal: *get general information about the owner*

1 Name […]
2 Gender [m/f]
3 Birth year [...]
4 Email [...]
5 Telephone number [...]
6 Hometown [...]
7 Native tongue: [dutch/dialect+.../other+...]
8 Through which dog pension/school did you receive this questionnaire? [list]
9 Are you the dog's primary caretaker? [y/n/we share responsibility]
    9.1 If not/shared:
        9.1.1 What is the gender of the (other) caretaker? [m/f]
        9.1.2 What is the birth year of the (other) caretaker? [...]
        9.1.3 What is the native tongue of the (other) caretaker? [dutch/dialect+.../other+...]

10. The dog's environment/ family

Goal: Understand more about the (social) environment of the dog

1 How many adults are there in your household in total? [number]
2 How many children (<18 years old) are there in your household?
3 Are any of the children younger than 4 years old? [y/n]
4 How many dogs have you had before this dog? [0 – >10]
5 Do you currently have additional dogs in your household? [y/n]
    5.1 If so, how many and what breed? (ex: 2 Beagles, 1 Border Collie)
6 Do you also have other pets (beside dogs) that your dog interacts with? [y/n]
    6.1 If so, how many and what kind? (ex. 3 cats, 2 horses, 1 bunny) [...]

11. Questions & Remarks:

If you have anything else you would like to add or ask you can do it here: [...]

Note:

The online (Dutch version) of the questionnaire can be found here:
https://spreadsheets.google.com/spreadsheet/viewform?formkey=dFB1YWxoaXlZbXBUeDFLTHZiOWF2d1E6MQ
(also accessible through the following website: http://hondentaal.blogspot.com)

Appendix B

Questionnaire score calculation

Age at which the dog was acquired:
Given that the scores I gathered from question 1.6 were concentrated around two extremes I could easily divide them in the following two categories: puppy (*min* = 0 weeks *max* = 10 weeks) and adult (*min* = 1 year, *max* = 3 years).

Playfulness, curiosity, intelligence, sociability:
Directly measured on a scale from 0 to 5 (not at all – a lot) on question 3.

Human-orientedness:
Obtained by summing up the scores from question 4.

Time spent with people:
Calculated by multiplying the number of hours the dogs spends actively involved with people on a daily basis (question 5.1.1) times two and adding it to the number of hours the dog spends daily in the proximity of people (question 5.1.2).

Time spent alone:
Directly obtained from question 5.1.4.

Time owner actively engages with dog:
The number of hours the owner spends with his/her dog per week were divided into 11 categories (0–5, 6–10, 11–15, 16–20 etc until 'more than 50") ranked 1– 11. (question 6.7)

Time owner actively engages with dogs <u>and</u> talks:
A combined score of how much the owner talks to the dog (question 6.8, scale 0 to 7) and how much time he/she spends with her dog (ranks of question 6.7), calculated by multiplying the first with the second.

Lexicon size:
Obtained by summing up input from questions 7.1.1, 7.2.1 and 7.3.1.

Quality of the relationship:
Calculated by summing up the input from questions 8.1, 8.2, 8.4 and 8.5, and then subtracting the input from question 8.3.

Medication:
Dogs were split into two categories: those who were under medication (due to health problems) and those who were not. No distinction was made based on the type of illness or medication.

Training:
Dogs were split into two categories: those who had received professional training and those who had not. No distinction was made based on the level or type of training.

Appendix C:

Raw experiment data

Participant ID	Breed	Age* (years; months)
DOG05	Labrador	3;5
DOG14	Boxer	8;0
DOG15	Kooikershond	1;1
RIJ04	Duitse Staande	.
RIJ05	Jack Russel	.
RIJ07	Franse Bulldog	7;2
RIJ08	Franse Bulldog	9;1
RIJ09	mix: Chihuahua	3;11
RIJ10	Duitse Herder	.
TAS01	Duitse Herder	1;3
TAS02	Shih – tzu	2;7
TAS04	Golden Retriever	3;9
TAS05	Rhodesian Ridgeback	.
TAS06	Rhodesian Ridgeback	3;3
TAS07	Airdale Terrier	9;10
TAS08	Fox Terrier	.
TAS09	Labrador Retriever	2;8

*age at moment of testing

Abbreviations (see next page)
FAM: familiarization
excl. #1: excluding first trial
Order: 0 = *new* trial first, 1= *old* trial first
FA: *faap* (new) trials
FE: *feep* (old) trials
|FA-FE|: absolute value of the difference between new and old trials (mean LT new trials **minus** mean LT old trials)
Puppet: 0 = without puppet during test trial, 1 = with puppet during test trials
dot: missing date or not applicable.
Note: All LT scores are the **mean LT from the two observers**.

ID	Gender	# trials	order	puppet	LT_FAM	LT_FA_ mean	LT_FE_ mean	LT_ALL_ mean	LT_ puppet	LT_no_ puppet	LT_FA_mean exlc #1	LT_FE_mean excl #1	LT_#1	LT_#2	LT_#3	LT_#4	LT_#5	LT_#6	LT_#7	LT_|FA-FE|
RIJ04	f	2	0	1	85282	8800	2289	5544	5544	.	.	.	8800	2289	.	.	.	.	.	6512
RIJ05	m	2	0	1	35547	5049	1452	3250	3250	.	.	.	5049	1452	.	.	.	.	.	3597
RIJ07	f	2	0	1	38421	7628	1435	4531	4531	.	.	.	7628	1435	.	.	.	.	.	6192
RIJ08	f	4	0	1	12400	5872	1975	3923	3923	.	945	1975	10798	3046	945	904	.	.	.	3897
RIJ09	f	4	0	1	15639	2893	1947	2420	2420	.	1399	1947	4388	2580	1399	1315	.	.	.	946
RIJ10	m	4	0	1	18562	1298	612	955	955	.	1309	612	1287	651	1309	574	.	.	.	685
TAS01	f	4	0	0	10629	6663	0	3332	.	3332	581	0	12746	0	581	0	.	.	.	6663
TAS02	f	4	0	0	66078	3738	728	2233	.	2233	7477	728	0	0	7477	1456	.	.	.	3010
TAS04	m	4	0	0	15907	7186	2007	4596	.	4596	8624	2007	5749	.	8624	220	.	.	.	5179
TAS05	f	4	0	0	28731	8229	2699	6386	.	6386	6693	2699	9766	.	6693	2699	.	.	.	5531
TAS06	f	4	0	0	56585	6235	906	3570	.	3570	7577	906	4893	718	7577	1094	.	.	.	5329
TAS07	m	4	0	1	29489	1766	1286	1526	1526	.	1547	1286	1986	1847	1547	725	.	.	.	480
TAS08	m	4	0	1	30739	2266	1428	1847	1847	.	1065	1428	3467	2276	1065	580	.	.	.	838
TAS09	f	4	0	1	12341	2403	2992	2698	2698	.	2209	2992	2597	3703	2209	2282	.	.	.	589
DOG05	m	2	0	0	49942	5465	2647	4056	.	4056	.	.	5465	2647	.	.	.	.	.	2818
DOG14	f	7	1	0	75365	2209	0	947	.	947	2209	0	0	0	0	932	0	0	5694	2209
DOG15	f	7	1	0	28297	1979	0	989	.	989	1979	0	0	3957	0	0	0	0	0	1979

Conditions per participant.

ID	Fam.* length (min)	Puppet		No. of trials	Fixed trial duration	First trial type	Full test sequence
		Fam*	Test				
RIJ04	3	-	+	2	-	new	*new–old*
RIJ05	3	-	+	2	-	new	*new–old*
RIJ07	2.5	-	+	2	-	new	*new–old*
RIJ08	2.5	-	+	4	-	new	*new–old–new–old*
RIJ09	2.5	-	+	4	-	new	*new–old–new–old*
RIJ10	2.5	-	+	4	-	new	*new–old–new–old*
TAS01	3.5	-	-	4	-	new	*new–old–new–old*
TAS02	2.5	-	-	4	-	new	*new–old–new–old*
TAS04	2.5	-	-	4	-	new	*new–old–new–old*
TAS05	2.5	-	-	4	-	new	*new–old–new–old*
TAS06	2.5	-	-	4	-	new	*new–old–new–old*
TAS07	2.5	-	+	4	-	new	*new–old–new–old*
TAS08	2.5	-	+	4	-	new	*new–old–new–old*
TAS09	2.5	-	+	4	-	new	*new–old–new–old*
DOG05	2.5	+	-	2	+	old	*old–new*
DOG14	2.5	+	-	7	+	old	*old–new–old–new–old–old–new*
DOG15	2.5	+	-	7	+	old	*old–new–old–new–old–old–new*

* *Fam : Familiarization*