Anim Cogn (2015) 18:405–421
DOI 10.1007/s10071-014-0811-7

ORIGINAL PAPER

Comparing supervised learning methods for classifying sex, age, context and individual Mudi dogs from barking

Ana Larrañaga · Concha Bielza · Péter Pongrácz · Tamás Faragó · Anna Bálint · Pedro Larrañaga

Received: 5 June 2013 / Revised: 22 September 2014 / Accepted: 22 September 2014 / Published online: 12 October 2014

Abstract Barking is perhaps the most characteristic form of vocalization in dogs; however, very little is known about its role in the intraspecific communication of this species. Besides the obvious need for ethological research, both in the field and in the laboratory, the possible information content of barks can also be explored by computerized acoustic analyses. This study compares four different supervised learning methods (naive Bayes, classification trees, *k*-nearest neighbors and logistic regression) combined with three strategies for selecting variables (all variables, filter and wrapper feature subset selections) to classify Mudi dogs by sex, age, context and individual from their barks. The classification accuracy of the models obtained was estimated by means of *K*-fold cross-validation. Percentages of correct classifications were 85.13 % for determining sex, 80.25 % for predicting age (recodified as young, adult and old), 55.50 % for classifying contexts (seven situations) and 67.63 % for recognizing individuals (8 dogs), so the results are encouraging. The best-performing method was *k*-nearest neighbors following a wrapper feature selection approach. The results for classifying contexts and recognizing individual dogs were better with this method than they were for other approaches reported in the specialized literature. This is the first time that the sex and age of domestic dogs have been predicted with the help of sound analysis. This study shows that dog barks carry ample information regarding the caller's indexical features. Our computerized analysis provides indirect proof that barks may serve as an important source of information for dogs as well.

Keywords Mudi dog barks · Acoustic communication · Feature subset selection · Machine learning · Supervised classification · *K*-fold cross-validation

Introduction

Canine communication (including dog–human communication) has become a well-studied topic among ethologists in the last decade. Most efforts have focused on how and to what extent dogs are able to understand different forms of human communication, through visual gestures (Reid 2009), voice recognition (Adachi et al. 2007), acoustic signals for ceasing or intensifying their activity (McConnell and Baylis 1985; McConnell 1990), and ostensive signals (Téglás et al. 2012). However, it has also been found that dogs can get their message across to humans, for example, by turning their head or alternating their gaze between the human and their target (Miklósi et al. 2000), and that dogs can emulate other behavioral forms so as to convey feelings, of guilt for example, in an appropriate situation (Hecht et al. 2012).

Unlike taxon-specific chemical and visual communication (Meints et al. 2010; Wan et al. 2012), acoustic signals

Electronic supplementary material The online version of this article (doi:10.1007/s10071-014-0811-7) contains supplementary material, which is available to authorized users.

A. Larrañaga
Student at the Universidad Alfonso X El Sabio, Av. Universidad, 1, 28691 Villanueva de la Cañada, Madrid, Spain

C. Bielza · P. Larrañaga (✉)
Computational Intelligence Group, Universidad Politecnica de Madrid, Campus de Montegancedo, 28660 Boadilla del Monte, Madrid, Spain
e-mail: pedro.larranaga@fi.upm.es

P. Pongrácz · T. Faragó · A. Bálint
Department of Ethology, Biological Institute, Eötvös Loránd University, 1117 Pázmány Péter sétány 1/c, Budapest, Hungary

are regarded as highly conservative and uniformly constructed within such broad groups of animals as avian and mammalian species. Morton (1977) provided a set of so-called motivation-structural rules to explain this point. According to his theory, the quality of the sound (pitch, tonality) strongly depends on the physical (anatomical) constraints of the animal's voice-producing tract, which in turn depends on the physical features of the animal itself (size, for example). Stronger, larger specimens within a species will usually be the dominant, aggressive animals and smaller, younger individuals are usually the subordinates. Thus, the typical vocalizations (low pitched, broadband, noisy) emitted by the larger, more aggressive individuals, for example, could, according to Morton, evolve into the trademarks of agonistic inner states. Similarly, the typical vocal features of a smaller, subordinate animal (high pitched, narrow band, tonal) could project the lack of aggressive intent communicative meaning.

Dogs have a rich vocal repertoire, see (Cohen and Fox 1976; Tembrock 1976; Yeon 2007), like other closely related wild members of the Canidae family. The ethological analysis of the possible functions of canine vocalizations has so far provided data about the individual-specific content of wolf howls (Mazzini et al. 2013; Root-Gutteridge et al. 2013), the indexical content of dog growls, related to the caller's body size (Taylor et al. 2008, 2010; Faragó et al. 2010a; Bálint et al. 2013), and the context-specific content of dog growls (Faragó et al. 2010b; Taylor et al. 2009). However, even though barking is considered to be the most characteristic form of dog vocalization, exceeding the barks of wolves and coyotes both in its frequency of occurrence and variability (Cohen and Fox 1976), the functional aspects of dog barks are surprisingly little known. The theoretical framework for the information content and evolution of barking in the dog involves very different assumptions, ranging from the theory that it is a non-communicative byproduct of domestication (Coppinger and Feinstein 1991), through the low-information level mobbing signal theory (Lord et al. 2000), to the context-specific information source theory (Feddersen-Petersen 2000; Yin 2002; Pongrácz et al. 2010). As dogs are the oldest domesticated companions of humans (Druzhkova et al. 2013), dog barking may have acquired a 'new target audience' in humans during the many 1,000 years of coexistence. A possible indirect proof of this is a series of playback experiments which showed that humans are able to correctly categorize barks according to their contexts (Pongrácz et al. 2005). As for contextual content, human listeners also had consistent opinions about the inner state of the barking dogs, and the acoustic analysis of the barks revealed that humans base their decision on the kinds of acoustic parameters of the barks that were expected on the basis of Morton's theory (Pongrácz et al. 2006). Besides the pitch and the harmonic-to-noise ratio, however, it was found that the inter-bark interval (or 'pulsing') of the barks is also important when assessing the inner state of the barking dog.

Although there are convincing empirical demonstrations that dog barks show acoustic features that are seemingly context specific (Yin 2002; Pongrácz et al. 2005), and we have also learned that humans can decipher information from dog barks regarding the context of vocalization and the inner state of the animal, it is less well understood whether dog barks carry an equally rich (or even richer) content of information for another dog. Until now, there have been only a few experiments with dogs as subjects which revealed that dog barks do carry individual-specific cues. One used a habituation–dishabituation paradigm (Maros et al. 2008; Molnár et al. 2009), and the other was a computerized bark analysis study (Molnár et al. 2008). These results raise the question of whether dog barks carry a much wider set of information about the vocalizing animal than humans are able to decipher. Another intriguing problem is which acoustic parameters could be responsible for the finer details of the information content of dog barks. Based on the vast literature of vocalization-based sex and individual recognition in other species, e.g., African wild dog, *Lycaon pictus* (Hartwig 2005); white-faced whistling duck, *Dendrocygna viduata* (Volodin et al. 2005); or Wied's black-tufted-ear marmosets, *Callithrix kuhlii* (Smith et al. 2009), one might expect dog barks to also carry specific cues of the caller's individual features, such as sex and age, for example. There are, however, considerable obstacles in testing such subtle pieces of information using classical techniques (i.e., playback). Fortunately, the current age of computer-based methods opens up the possibility for analyzing and testing lots of sound samples with the help of artificial intelligence.

Machine learning techniques have been used in behavioral research on acoustic signals for a wide range of species, see Table 1. For dolphins, artificial neural networks have been applied to model dolphin sonar, specifically for discriminating differences in the wall thickness of cylinders using time and frequency information from the echoes (Au et al. 1995). Also, support vector machines and quadratic discriminant function analysis have been used to classify fish species according to their echoes using a dolphin-emulating sonar system (Yovel and Au 2010), and Gaussian mixture models and support vector machines have been employed to classify echolocation clicks from three species of odontocetes (Roch et al. 2008). Differentiation of categories or graded barks in mother-calf vocal communication in Atlantic walrus have been analyzed with artificial neural networks and discriminant functions (Charrier et al. 2010). Frog song identification to recognize frog species has been carried out with *k*-nearest neighbor classifiers and support vector machines (Hunag et al. 2009). Linear discriminant analysis, decision tree and support

Table 1 Examples of machine learning technique usage from acoustic signals for different species with different aims

Animal	Aim	Technique	Reference
Dolphin	Discriminate cylinder thickness	ANN	Au et al. (1995)
	Classify fish species	SVM, quadratic DFA	Yovel and Au (2010)
Odontocete	Classify echolocation clicks	GMM, SVM	Roch et al. (2008)
Walrus	Classify barks in mother-calf communication	ANN, DFA	Charrier et al. (2010)
Frog	Classify species	kNN, SVM	Hunag et al. (2009)
		Linear DFA, trees, SVM	Acevedo et al. (2009)
Bird	Classify species	Linear DFA, trees, SVM	Acevedo et al. (2009)
	Recognize individuals	GMM	Cheng et al. (2010)
Bat	Classify species	ANN, DFA	Parsons (2001), Parsons and Jones (2000)
		Trees	Adams et al. (2010)
		Random forests, SVM	Armitage and Ober (2010)
		ANN, DFA, kNN	Britzke et al. (2011)
Cricket, grasshopper	Classify species	ANN	Chesmore (2001)
Marmot	Classify identity, age and sex	DFA	Blumstein and Munos (2005)
Suricate	Predict predator type	DFA	Manser et al. (2002)
African elephant	Classify vocalization type	HMM	Clemins (2005)
	Classify contexts	HMM	Clemins (2005)
	Recognize individuals	HMM	Clemins (2005)
Female elephant	Classify rumbles by oestrous cycle phase	HMM	Clemins (2005)
Artic fox	Recognize individuals	DFA	Frommolt et al. (2003)
African wild dog	Recognize individuals	DFA	Hartwig (2005)
Domestic dog	Classify contexts	DFA	Yin and McCowan (2004)
	Recognize individuals (breeds)	DFA	Yin and McCowan (2004)
Mudi dog	Classify contexts	Gaussian NB	Molnár et al. (2008)
	Recognize individuals	Gaussian NB	Molnár et al. (2008)

ANN artificial neural network, *SVM* support vector machine, *DFA* discriminant function analysis, *GMM* Gaussian mixture model, *k*NN *k*-nearest neighbor classifier, *HMM* hidden Markov model, *NB* naive Bayes

vector machines have been employed to automate the classification of calls of several frog and bird species (Acevedo et al. 2009). Gaussian mixture models have also been used for individual animal recognition in birds (Cheng et al. 2010). Bat species have been acoustically identified using artificial neural networks (Parsons 2001; Britzke et al. 2011), discriminant function analysis (Parsons and Jones 2000; Britzke et al. 2011), classification trees (Adams et al. 2010), *k*-nearest neighbors (Britzke et al. 2011) as well as other classifiers (random forests and support vector machines) whose behavior has been compared (Armitage and Ober 2010). Artificial neural networks have been used to discriminate between the sounds of different animals within a group of British insect species (Orthoptera), including crickets and grasshoppers (Chesmore 2001). Blumstein and Munos (2005) found potentially significant information about identity, age and sex encoded in yellow-bellied marmots calls using discriminant function analysis. For suricates, discriminant function analysis was chosen to predict the predator type (mammal, bird and snake) from the alarm calls (Manser et al. 2002). Hidden Markov models have been used to analyze African elephant vocalizations and speaker identification, discrimination of rumbles in different contexts, and oestrous cycle phase determination from rumbles of female elephants (Clemins 2005). Moreover, other work has focused on identifying calls from different animals such as bears, eagles, elephants, gorillas, lions and wolves, with *k*-nearest neighbor classifiers, artificial neural networks and hybrid methods (Gunasekaran and Revathy 2011).

For canids, research analyzing the acoustic measures of barks with machine learning methods is limited, see Table 1. Discriminant functions have been used for individual recognition within a wild population of Arctic foxes (Frommolt et al. 2003) and African wild dogs (Hartwig 2005). Domestic dog barks have been analyzed again using discriminant analysis (Yin and McCowan 2004) for classification into context-based subtypes (three different contexts) and in order to identify individual dogs. These two tasks were further refined in the same paper to categorize each individual's barks into separate contexts and identify the individual barking within each context. A total of 4,672 barks were recorded from ten dogs of six different breeds, and 120 variables were extracted from the spectrograms. More recently, 6,006 barks of 14 Mudi breed individuals were recorded under six different communicative situations (Molnár et al. 2008). After processing the spectrograms of their signals, a genetic programming-based heuristic guided the construction of new descriptors. The aims were the same as in Yin and McCowan (2004), although the machine learning technique was a Gaussian naive Bayes classifier.

In this paper, we extend Molnár et al.'s research in several ways. As in Molnár et al. (2008), we classify barks into contexts and identify individual barks. Unlike Molnár et al., we also investigate whether barks encode information about dog sex and age. Also, we specify context classification per individual dog and recognize individual bark per context. Therefore, we have six different classification problems concerning sex, age, contexts, contexts per individual, individuals and individuals per context. Moreover, for each of these six problems, a thorough set of four machine learning models (Gaussian naive Bayes, classification trees, *k*-nearest neighbors and logistic regression) are trained from a database of 800 barks corresponding to 8 Mudi dogs in seven behavioral contexts. Their performance is estimated using cross-validation (*K*-fold scheme) which assesses the ability to classify barks that had not been previously encountered. Given an incoming Mudi dog bark, two models (Gaussian naive Bayes and logistic regression) output the probability of each class value, whereas the other two models deterministically provide the predicted class value. Gaussian naive Bayes assumes normality and independence of the features given the class value. Logistic regression uses the sigmoid function of a linear combination of the features as the probability of each class value. Classification trees hierarchically partition the feature space. Finally, *k*-nearest neighbors simply predicts the class value by majority voting in a feature space neighborhood. The diversity of these four models is representative of the available supervised classifiers. Rather than using all the extracted acoustic measures, we selected relevant features with two methods, filter and wrapper, for each machine learning model. Whereas wrapper methods use a predictive model to score feature subsets, filter methods use a proxy measure instead of the classification accuracy to score the selected features.

Methods

Subjects

Barks recorded from Mudi dogs were used for this study. The Mudi is a medium-sized Hungarian herding dog breed. The Mudi breed standard is listed as #238 with the FCI (*Fédération Cynologique Internationale*). Initially, we collected 7,310 barks from 27 individuals. The number of barks per dog ranged from 8 to 1,696. These barks were recorded in different number of bouts for each dog. Trying to minimize the effect of pseudoreplication, we only considered dogs whose initial number of barks was greater than 300. From each of these 8 dogs, 100 barks were randomly selected using a systematic sampling procedure, thereby balancing the number of samples coming from each individual. Table 2 contains the characteristics of these selected 800 barks according to sex ratio (male–female 3:5), age (ranging from 1 to 10 years old), number of bouts for each dog (with a minimum of 5 and a maximum of 14) and number of barks per dog in each of the seven contexts. Age values are grouped into intervals to form a three-valued class variable: young dogs (1–3 years old), adult dogs (4–8 years old) and old dogs (more than 8 years old).

Recording and processing of the sound material

Recording contexts

Recordings were made using a Sony TCD-100 DAT tape recorder and Sony ECM-MS907 microphone on Sony PDP-65C DAT tapes. During recording of the barks, the experimenter held the microphone at a distance of 3 to 4 m from the dog. We collected bark recordings in seven different behavioral contexts. With the exception of two contexts (Alone and Fight), all recordings were done at the dog's residence. Barks of the Fight context were recorded at dog training schools. The training school dogs were also taken to a park or other suitable outdoor area to record the Alone barks. The seven situations are as follows:

- Alone ($N = 106$ recordings): The owner and the experimenter (male, 23 years old) took the dog to a park or other outdoor area, where the dog was tied to a tree or fence by its leash. The owner left the dog and walked out of the dog's sight, while the experimenter remained with the dog and recorded its barks.
- Ball ($N = 131$): The owner held a ball (or one of the dog's favorite toys) approximately 1.5 m in front of the dog.
- Fight ($N = 131$): For dogs to perform in this situation, the trainer acts as if he intends to attack the dog–owner dyad. Dogs are expected to bark aggressively and even bite the trainer's glove. The owner keeps the dog on a leash during this exercise.
- Food ($N = 106$): The owner held the dog's food bowl approximately 1.5 m in front of the dog.
- Play ($N = 89$): The owner was asked to play a game with the dog, such as tug-of-war, chasing or wrestling.

Table 2 Characteristics of the bark data set with seven context categories: Alone, Ball, Fight, Food, Play, Stranger and Walk

Context												
Number	Dog	Sex	Age (years)	Bouts	Alone	Ball	Fight	Food	Play	Stranger	Walk	Total
1	Bogyó	Male	1	5					50	50		100
2	Derüs	Female	2	15				50		50		100
3	Fecske	Female	2	10	25	25	25			25		100
4	Guba	Female	5	14	50	50						100
5	Harmat	Female	4	7			50			50		100
6	Sába	Female	6	7		25	25	25	25			100
7	Ügyes	Male	10	6	17	17	17	17		17	17	102
8	Merse	Male	7	6	14	14	14	14	14	14	14	98
Total					106	131	131	106	89	206	31	800

The experimenter recorded the barks emitted during this interaction.

- Stranger ($N = 206$): The experimenter acted as the 'stranger' for all the dogs and appeared at the dog owners' garden or front door. The experimenter recorded the barking dog for 2–3 min. The owner was not in the vicinity (in the garden, or near to the entrance) during the recording.
- Walk ($N = 31$): We asked the owner to behave as if he/she was preparing to go for a walk with the dog. For example, the owner took the dog's leash in her/his hand and told the dog 'We are leaving now.'

Initial processing of the sound material

The recorded material was digitalized with a 16-bit quantization and 44.10 kHz sampling rate using a TerraTec DMX 6Wre 24/96 sound card. As each recording could contain at least three or four barks, individual bark sounds were manually segmented and extracted. This process resulted in a final collection of 7,310 sound files containing only a single bark sound. Obviously, these sounds are not independent from a statistical point of view. As some of the machine learning methods used in this work assume that the samples are independent and identically distributed, we randomly selected non-consecutive barks, alleviating in this way the pseudoreplication effect. The final data set contains 800 barks from the initial 7,310 sound files.

Sound analysis

Based on the initial parameter set used in Molnár et al. (2008), 29 acoustic measures were extracted from the bark samples with an automated Praat script, see Table 3 and Fig. 1.

The energy, loudness and the long-term average spectrum (LTAS) are measurements of sound energy, and the LTAS parameters reflect its change over time, whereas the spectral parameters show the distribution of energy over the frequency components.

According to the source–filter framework (Fant 1976), the fundamental frequency is the lowest harmonic component of the source signal that is produced in the larynx by the movements of the vocal fold. Measurements of the fundamental show the modulation of this source signal over time. One voice cycle is the unit of the movements of the vocal folds. During sound production, the repeated opening and closing of the vocal folds generates cyclic pressure changes in the exhaled air, which will be the sound wave itself. Measurements of the vocal cycles show the regularities in voice production.

Finally, tonality or harmonics-to-noise ratio gives the proportion of regular, tonal frequency components over the noise caused by the irregular movements of the vocal folds, or the turbulences in the air flow in the vocal tract. These measurements are capable of describing the quality of the sound and its change over time.

The process is illustrated in Fig. 2 (top).

Supervised classification

A common machine learning task is pattern recognition (Duda et al. 2001), in which two different problems are considered depending on the available information. We always started from a data set in which each case or instance (a single bark sound in this paper) is characterized by features or variables (29 acoustical measures in our case). In a supervised classification problem, an additional variable—called the class variable—contains the instance label (sex, age, context or individual in this paper), and we look for a model able to predict the label of a new case with known features. Alternatively, in an unsupervised classification problem or clustering (Jain et al. 1999), the label is missing and the aim is to form groups or clusters with cases

Table 3 Twenty-nine acoustic measures extracted from barking recordings

Name	Description	Variable
Measurements of sound energy		
Energy	Amount of energy in the sound ($\mathrm{Pa}^2 \cdot s$)	X_1
Loudness	Loudness	X_{10}
Ltasm	Mean long-term average spectrum (ltas)	X_{23}
Ltass	Slope of the ltas	X_{24}
Ltasp	Local peak height between 1,700 and 3,200 in the ltas	X_{25}
Ltasd	Standard deviation of the ltas	X_{26}
Measurements of spectral energy		
Banddensity	Density of the spectrum between 2,000 and 4,000 Hz	X_2
Centerofgravityfreq	Average frequency in the spectrum	X_3
Deviationfreq	Standard deviation of the frequency in the spectrum	X_4
Skewness	Skewness of the spectrum	X_5
Kurtosis	Kurtosis of the spectrum	X_6
Cmoment	Non-normalized skewness of the spectrum	X_7
Energydiff	Energy difference between 0–2,000 and 2,000–6,000 Hz bands	X_8
Densitydiff	Density difference between 0–2,000 and 2,000–6,000 Hz bands	X_9
Measurements of the source signal		
Pitchm	Mean fundamental frequency (F0) in Hertz	X_{11}
Pitchmin	Minimum F0	X_{12}
Pitchmax	Maximum F0	X_{13}
Pitchmint	Time point of the minimum F0 (s)	X_{14}
Pitchmaxt	Time point of the maximum F0 (s)	X_{15}
Pitchd	Standard deviation of the F0	X_{16}
Pitchq	Lower interquantile of the F0	X_{17}
Pitchslope	Mean absolute slope of the F0	X_{18}
Pitchslopenojump	Mean slope of the F0 without octave jump	X_{19}
Measurements of the voice cycles		
Ppp	Number of voice cycles	X_{20}
Ppm	Mean number of voice cycles	X_{21}
Ppj	Jitter	X_{22}
Measures of the tonality		
Harmmax	Maximum tonality	X_{27}
Harmmean	Mean tonality	X_{28}
Harmdev	Standard deviation of the tonality	X_{29}

(dog barks) that are similar with respect to the features at hand.

In this paper, we apply supervised classification methods to automatically learn models from data. These models will be used to separately predict dog sex, dog age, context and the individual dog from a set of predictor variables capturing the acoustical measures of the dog barks.

In a binary supervised classification problem, there is a feature vector $\mathbf{X} \in \mathbb{R}^{\mathbf{n}}$ whose components, $X_1, \ldots, X_n$, are called predictor variables, and there is also a label or class variable C taking values on $\{0, 1\}$. The task is to induce classifier models from training data, which consists of a set of N observations $\mathcal{D}_N = \{(\mathbf{x}^{(1)}, \mathrm{c}^{(1)}), \ldots, (\mathbf{x}^{(\mathrm{N})}, \mathrm{c}^{(\mathrm{N})})\}$ drawn from the joint probability distribution $p(\mathbf{x}, \mathrm{c})$, see Table 4. In our dog data set, $n = 8$ acoustical measures and $N = 800$ bark sounds. The classification model will be used to assign labels to new instances, $\mathbf{x}^{(\mathrm{N}+1)}$, only characterized by the values of the predictor variables.

To quantify the goodness of a binary classification model, true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are counted over the test data and placed in a confusion matrix. This confusion matrix contains in its diagonal the TP and TN observations. Then, we can define the error rate as $\frac{|FN|+|FP|}{N}$, where $N = |TP| + |FP| + |TN| + |FN|$ is the total number of instances, or equivalently, the accuracy as $\frac{|TP|+|TN|}{N}$.

Dog sex classification is binary, $\Omega_C = \{\text{Female, Male}\}$, where there are two possible errors: predict a Male as a Female dog, and alternatively predict a Female as a Male.

The other classifications are multiclass, where C takes $r > 2$ possible class values. Let $\Omega_C = \{1, 2, \ldots, r\}$ denote this set. Thus, $\Omega_C = \{\text{Young, Adult, Old}\}$ for age, $\Omega_C = \{\text{Alone, Ball, Fight, Food, Play, Stranger, Walk}\}$ for contexts, and $\Omega_C = \{\text{dog1,...,dog8}\}$ for individuals in our case. The $r \times r$-dimensional confusion matrix contains all pairwise counts, m_{ij}, the number of cases out of N from the real class c_i classified by the model as c_j. The accuracy is given by $\sum_{i=1}^{r} m_{ii}/N$.

Accuracy estimation of supervised classification models

An important issue is how to honestly estimate the (expected) accuracy of a classification model when using this model for classifying unseen (new) instances. A simple method is to partition the whole data set into two subsets: the training subset and the test subset. According to this training and testing scheme, the classification model is learned from the training subset, and it is then used in the test subset for the purpose of estimating its accuracy. However, the information in the data set is under-used, as the classification model is learned from a subset of the original data set.

In this paper, we will use an estimation method called K-fold cross-validation (Stone 1974). This uses the whole data set to honestly learn the model. The data set is partitioned into K folds of approximately the same size. Each

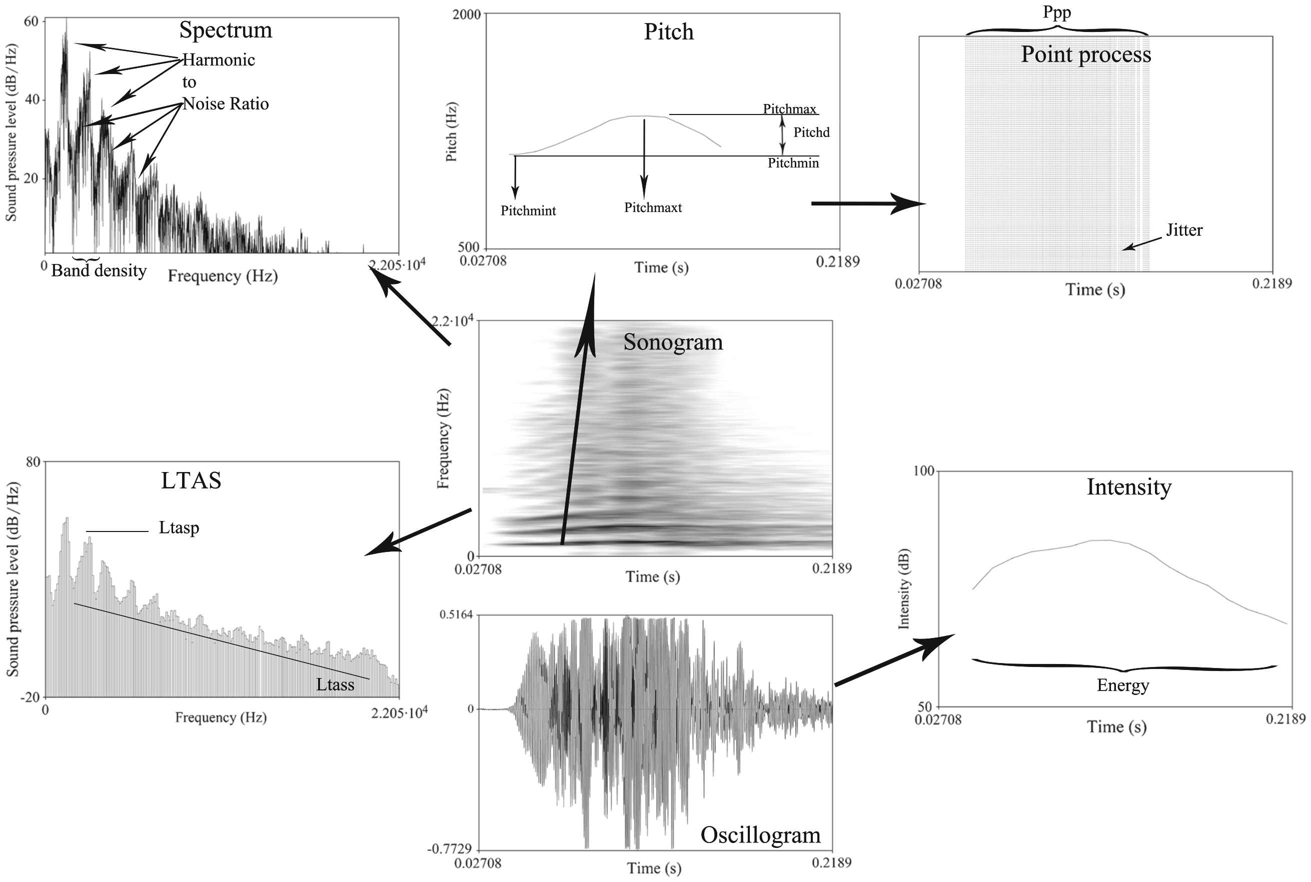

Fig. 1 Main parameters measured for the acoustic analysis using Praat functions. The oscillogram shows the actual complex waveform of a single bark. The amplitude of the waveform shows the intensity change over time, which is represented here as the intensity profile. The energy parameter is the overall energy transferred by the sound over time. Fast Fourier transformation is used to create a sonogram which shows the frequency spectrum of the bark over time. Autocorrelation method was applied to extract the fundamental frequency and its profile depicted as the pitch object. The fundamental frequency is the frequency of opening and closing cycles of the vocal fold, which is represented by the point process object where every vertical line represents one vocal cycle. This can be used to measure the periodicity of the sound and irregularities in sound production (jitter). The spectrum shows the overall power of each frequency component. The harmonic-to-noise ratio gives the ratio of harmonic spectral components (the upper harmonics of the fundamental frequency) over the irregular, noisy components. Finally, the long-term average spectrum (LTAS) represents the average energy distribution over the frequency spectrum

fold is left out of the learning process, which is carried out with the remaining $K-1$ folds, and used later as a test set. This process is repeated K times. Thus, every instance is in a test set exactly once and in a training set $K-1$ times. The model accuracy is estimated as the mean of the accuracies for each of the K test sets. In our experiments, we will fix the value of K to 10.

Feature subset selection

The feature subset selection (FSS) problem (Liu and Motoda 1998) refers to the question of whether all the n predictor features are really useful for classifying the instances with a given model. The FSS problem can be formulated as follows: Given a set of candidate features, select the best subset under some classification learning method.

This dimensionality reduction by means of an FSS process has several potential advantages for a supervised classification model, such as the reduction in the cost of data acquisition, an improved understanding of the final classification model, a faster induction of the classification model and an improvement in classifier accuracy.

FSS can be viewed as a search problem, where each state in the search space specifies a subset of selectable features. An exhaustive search of all possible feature subsets, given by 2^n, is usually unfeasible in practice because of the large computational burden, and heuristic search is usually used.

For a categorization of FSS, see Saeys et al. (2007). There are two main types of FSS depending on the function used to measure the goodness of each selected subset. In the wrapper approach to the FSS, the accuracy reported by

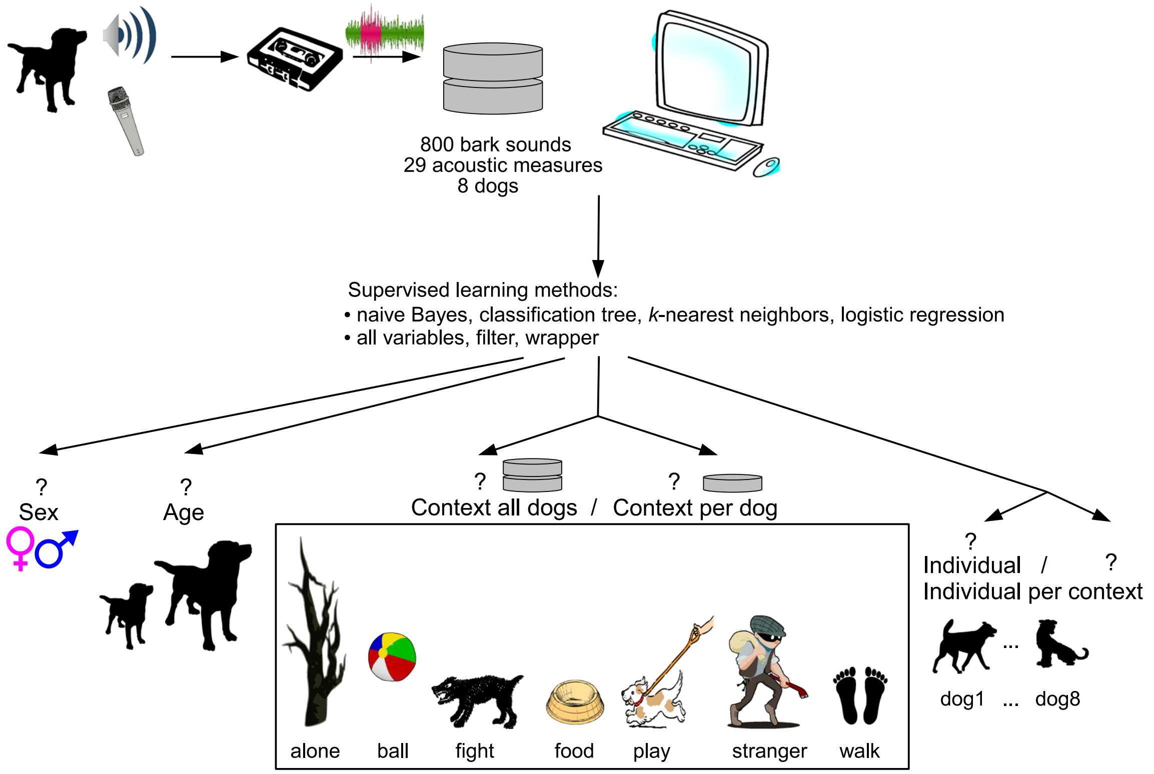

Fig. 2 Diagram of the study: data preprocessing (*top*) and questions to be answered by machine learning models (*bottom*)

Table 4 Raw data in a supervised classification problem: N denotes the number of labeled observations, each of them characterized by n predictor variables, $X_1, \ldots, X_n$ and the class variable C

	X_1	...	X_n	C
$(\mathbf{x}^{(1)}, \mathrm{c}^{(1)})$	$x_1^{(1)}$	...	$x_n^{(1)}$	$c^{(1)}$
$(\mathbf{x}^{(2)}, \mathrm{c}^{(2)})$	$x_1^{(2)}$	...	$x_n^{(2)}$	$c^{(2)}$
...		...		...
$(\mathbf{x}^{(\mathrm{N})}, \mathrm{c}^{(\mathrm{N})})$	$x_1^{(N)}$	...	$x_n^{(N)}$	$c^{(N)}$
$\mathbf{x}^{(\mathrm{N}+1)}$	$x_1^{(N+1)}$	...	$x_n^{(N+1)}$	?

$\mathbf{x}^{(\mathrm{N}+1)}$ denotes the new observation to be classified by the supervised classification model

a classifier guides the search for a good subset of features. We have used a greedy stepwise search in our experiments, i.e., one that progresses forward from the empty set selecting at each step the best option among adding a variable not yet included within the model and deleting a variable from the current model. The search is halted when neither of these options improves model accuracy. When the learning algorithm is not used in the evaluation function, the goodness of a feature subset can be assessed using only intrinsic data properties, such as an information theory based evaluation function. This is the filter approach to the FSS problem. In this paper, we apply both wrapper and filter approaches to the FSS problem. For the second type, a multivariate filter based on mutual information, called correlation feature selection, is used (Hall 1999). This tries both to minimize redundancy between selected features and maximize correlation with the class variable.

Supervised classification methods

Given an instance $\mathbf{x}$, supervised classification builds a function γ that assigns to $\mathbf{x}$ a class label in $\Omega_C = \{1, \ldots, r\}$. We provide a short description of each supervised classification method used.

Naive Bayes (Minsky 1961) is the simplest Bayesian classifier. A Bayesian classifier assigns the most probable *a posteriori* class to a given instance $\mathbf{x}$, i.e., it yields the c value of C that maximizes the posterior probability $p(c|\mathbf{x})$. Using the Bayes' theorem, this is equivalent to maximizing $p(c)p(\mathbf{x}|\mathrm{c})$. The naive Bayes is built upon the assumption of conditional independence of the predictive variables given the class. Computationally, this means that $p(\mathbf{x}|\mathrm{c})$ in the previous product is easily obtained as the product of all factors $p(x_j|c)$, $j = 1, \ldots, n$, each associated with one variable. The Gaussian naive Bayes classifier applies for continuous variables X_j following a Gaussian distribution f_j. Therefore, this model computes c such that

$$\max_{c\in\Omega_C} p(c)\prod_{j=1}^{n} f_j(x_j|c). \tag{1}$$

In a classification tree (Quinlan 1993), the learned function γ is represented by a decision tree. Each (non-leaf) node specifies a value test of some variable of the instance. Each descendant branch corresponds to one of the possible values for this variable. Each leaf node provides the class label given the values of the variables jointly represented by the path from the root to that leaf. Unseen instances are classified by sorting down the tree from the root to some leaf node testing the variable specified at each node. A classification tree is learned in a top-down manner (starting with the root node) by progressively splitting the training data set into smaller and smaller subsets based on variable value tests. This process is repeated on each derived subset in a recursive manner called recursive partitioning of the space representing the predictive variables. Key decisions are how to select which variable to test at each node in the tree, and how deep the tree should be, i.e., whether to stop splitting or select another variable and grow the tree further. These decisions make the differences between algorithms. The C4.5 algorithm used in this paper chooses variables by maximizing the gain ratio, which is the ratio of the information gain of X_j and C and the entropy of X_j, which are both concepts used in information theory. The algorithm incorporates post-pruning rules to avoid the tree becoming too deep thereby escaping from the training data overfitting, i.e., its failure to work well with new unseen instances.

The k-nearest neighbor classifier (Fix and Hodges 1951) is a nonparametric method that assigns to a given instance $\mathbf{x}$ the class label most frequently found among its k nearest instances; that is, the predicted class is decided by examining the labels of the k nearest neighbors and voting. A common distance used for obtaining the k nearest neighbors for a continuous variable $\mathbf{x}$ is the Euclidean distance. This classifier is a type of lazy learning where the function is only approximated locally and all computation is deferred until classification. In our experiments, we will fix $k = 1$.

Logistic regression (Le Cessie and van Houwelingen 1992), like naive Bayes, produces a posterior probability $p(c|\mathbf{x})$ for a given instance $\mathbf{x}$. For binary classification, the model assumes that it is a transformation of a linear combination of the input variables, given by

$$p(C = 1|\mathbf{x}) = 1/[1 + e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}],$$

where $\beta_0, \beta_1, \ldots, \beta_n$ are model parameters estimated from data by maximum likelihood. If $\mathcal{L}(\beta_0, \ldots, \beta_n)$ denotes the log-likelihood function of the data under this model, the problem is to find βs that maximize this function. The ridge logistic regression used in this paper adds a penalization term to $\mathcal{L}$, and the problem is then to maximize the function $\mathcal{L}(\beta_0, \ldots, \beta_n) - \lambda \sum_{j=1}^{n} \beta_j^2$, for βs where $\lambda > 0$ controls the amount of penalization. This penalty forces the parameters to shrink to zero achieving a reduction in the variance of the parameter estimates with an overall increased accuracy. For multiclass classification, the posterior probability of $c \neq r$ is given by

$$p(c|\mathbf{x}) = \frac{\mathrm{e}^{(\beta_0^{(\mathrm{c})}+\beta_1^{(\mathrm{c})}\mathrm{x}_1+\cdots+\beta_\mathrm{n}^{(\mathrm{c})}\mathrm{x}_\mathrm{n})}}{1+\sum_{l=1}^{r-1}\mathrm{e}^{(\beta_0^{(\mathrm{l})}+\beta_1^{(\mathrm{l})}\mathrm{x}_1+\cdots+\beta_\mathrm{n}^{(\mathrm{l})}\mathrm{x}_\mathrm{n})}}, \quad l = 1, \ldots, r-1 \tag{2}$$

and hence, $p(r|\mathbf{x})$ is derived from the others since they all sum to one. Note that in this multiclass case, we need a set of $n + 1$ parameters $\{\beta_0^{(l)}, \beta_1^{(l)}, \ldots, \beta_n^{(l)}\}$ for each l value, $l = 1, \ldots, r-1$; that is, a total of $(n+1)(r-1)$ parameters.

All the results were calculated using WEKA software (Hall et al. 2009).

Results

The six problems we will deal with are illustrated in Fig. 2 (bottom).

Sex

The k-nearest neighbor classifier produced the best results, with an accuracy of 85.13 %, with a wrapper feature selection (in bold), see Table 5. This model contains 12 predictor variables, see Table 16. The groups that record spectral energy and source signal variables are under-represented, according to the categorization of acoustic measures provided in Table 3.

For the female barks, the misclassification rate is 9.40 % (47 false males out of 500 real females), and it is higher for males, 24.00 % (72 false females from a total of 300 real males).

Table 6 shows the accuracies per dog of the k-nearest neighbor model with 12 predictors. The model accuracy when predicting the five female dogs is around 90 %, with the worst predictions for dog3 and dog4 (87.00 %), and the best for dog5 (97.00 %). The three male dogs are predicted with accuracies ranging from 73.00 % for dog1 to 79.41 % for dog7.

Supplementary Material contains the specifications of the best models for the prediction of the dog sex. For naive Bayes, the univariate conditional Gaussian densities for each predictor variable are shown. The structure of the classification tree model is also presented, as well as the

Table 5 Sex prediction

	All	Filter	Wrapper
Naive Bayes	71.00 %	71.13 %	77.13 %
Classification tree	78.13 %	72.75 %	81.50 %
k-Nearest neighbors	82.00 %	64.25 %	**85.13 %**
Logistic regression	76.88 %	70.50 %	78.63 %

Accuracies of the twelve models: three selection feature methods for each of the four supervised classifiers

Table 6 Sex prediction per dog

	Male 76.00 %	Female 90.60 %
Dog1	73.00 %	–
Dog2	–	90.00 %
Dog3	–	87.00 %
Dog4	–	87.00 %
Dog5	–	97.00 %
Dog6	–	92.00 %
Dog7	79.41 %	–
Dog8	75.51 %	–

Accuracies of the best model in Table 5 for each of the eight dogs. The overall accuracy of this model over the eight dogs is 85.13 %

coefficients of the logistic regression model. For the *k*-nearest neighbor classifier, the data set constitutes the model and therefore it is not shown.

Age

Table 7 (left) shows the age results. As for the sex prediction problem, *k*-nearest neighbors with a wrapper feature selection produced the best accuracy 80.25 %. The 15 selected variables in this model mainly contain measurements of spectral energy, sound energy and voice cycles. For this problem, the wrapper strategy outperformed the other strategies in the four supervised classification methods.

The confusion matrix in Table 7 (right) of the best model shows that a Young dog is classified as Old in only 2.67 % of cases (8 out of 300), while old dogs are misclassified as Young in 6.86 % of cases (7 out of 102). The error rates classifying Young, Adult and Old dogs are 21.00, 17.59 and 24.51 %, respectively. These figures suggest that it is easier to get it wrong when classifying Young and Old dogs.

Table 8 contains the accuracies per dog of the best model. This model provides a 79.00 % of accuracy when predicting Young dogs. This percentage is very similar for each of the three young dogs (dog1, dog2 and dog3). However, for the four adult dogs the model shows a wide range of accuracies, varying from 66.00 % (dog6) to 90.00 % (dog5). Dog7, that is the only old dog, is classified with an accuracy of 75.49 %.

Table 7 Age prediction

	All	Filter	Wrapper
Naive Bayes	68.50 %	65.63 %	71.88 %
Classification tree	70.88 %	69.13 %	74.13 %
k-Nearest neighbors	78.63 %	79.13 %	**80.25 %**
Logistic regression	75.63 %	73.88 %	76.00 %

Real class	Predicted class		
	Young	Adult	Old
Young	237	55	8
Adult	61	328	9
Old	7	18	77

Accuracies of the twelve models: three selection feature methods for each of the four supervised classifiers (top table). Confusion matrix of the best model: *k*-nearest neighbors wrapper (bottom table)

Table 8 Age prediction per dog

	Young 79.00 %	Adult 82.41 %	Old 75.49 %
Dog1	84.00 %	–	–
Dog2	74.00 %	–	–
Dog3	79.00 %	–	–
Dog4	–	85.00 %	–
Dog5	–	90.00 %	–
Dog6	–	66.00 %	–
Dog7	–	–	75.49 %
Dog8	–	88.77 %	–

Accuracies of the best model in Table 7 for each of the eight dogs. The overall accuracy of this model over the eight dogs is 80.25 %

Supplementary Material contains the specifications of the best models for the prediction of the dog age.

Context

A single model for all dogs. Table 9 (left) shows the results of a single model learned from the 800 barks to discriminate among the 7 contexts: Alone, Ball, Fight, Food, Play, Stranger and Walk.

k-nearest neighbor classifier and wrapper selection is once more the best-performing model with an accuracy of 55.50 %. The variables selected by this model correspond mainly to spectral energy and voice cycle measurements. Note that now we have a more difficult problem with more class values to be predicted (7 contexts) and consequently the estimated accuracy is expected to be lower.

From Table 9 (right), we can compute the contexts with the highest and lowest true positive rates that correspond to Fight (0.76) and Walk (0.35), respectively. The Ball

Table 9 Context prediction

	All	Filter	Wrapper
Naive Bayes	41.63 %	42.63 %	47.88 %
Classification tree	44.00 %	44.63 %	44.13 %
k-Nearest neighbors	50.88 %	50.75 %	**55.50 %**
Logistic regression	49.75 %	47.50 %	50.13 %

Real class	Predicted class						
	Alone	Ball	Fight	Food	Play	Stranger	Walk
Alone	46	15	7	17	6	14	1
Ball	11	64	5	22	5	23	1
Fight	8	4	100	3	4	11	1
Food	7	20	2	55	3	15	4
Play	8	8	2	10	44	11	6
Stranger	12	24	5	26	13	124	2
Walk	0	3	4	5	6	2	11

Accuracies of the twelve models: three selection feature methods for each of the four supervised classifiers (top table). Confusion matrix of the best model: *k*-nearest neighbors wrapper (bottom table)

context is often misclassified as Food and vice versa. The same holds for the Walk and Play pair. This is quite reasonable since both pairs define quite similar underlying concepts. Many barks under Fight or Alone situations are misclassified as Stranger. However, the Stranger context is usually confused with the Ball and Food context.

Table 10 contains the accuracies per dog of the best model. This model provides 43.40 % accuracy when predicting the Alone context, with extreme prediction accuracies for dog7 (52.94 %) and dog8 (14.29 %). The Ball context achieves 48.85 % accuracy, having dog7 and dog8 the worst (29.41 %) and best (64.29 %) predictions, respectively. These two dogs also present the worst and best predictions for the Food context. The model shows better accuracies for the Fight and Stranger contexts. In the Fight context, the 98.00 % of success for dog5 is noteworthy, whereas the worst behavior in the Stranger context is for dog7 (35.29 %). The Play and Walk contexts show highly variable accuracies for the different dogs.

Supplementary Material contains the specifications of the best models for the prediction of the dog context.

A model per dog. More refined dog-specific models are built here. By selecting instances from the same dog, the corresponding model will identify the context for that dog. A total of 96 models (8 dogs × 12 models per dog) have been considered, where only the performance of the best model is shown in Table 11.

Naive Bayes was the best model 3 times, *k*-nearest neighbors 4 times, and logistic regression in 2 cases. Regarding the feature subset selection methods, wrapper reports the best results for all 8 dogs.

Table 11 shows that accuracies decrease in proportion to the increase in the number of contexts. With two contexts, accuracies fall in the interval [78, 100 %]. The accuracies for the two dogs with four contexts are 74 and 73 %. Increasing the number of contexts to six and seven, the accuracies are 59.80 and 66.98 %, respectively.

Figure 3 displays, for the best models in Table 11, the mean number of selected variables by the five types of acoustic variables. Spectral energy and voice cycle measurements were the two groups with more often selected (in relative terms) variables regardless of the number of barks.

From the previous table, we select some models for the sake of illustration. Figure 4 shows the naive Bayes model which performed best for dog5, with only two observed contexts, Fight and Strange (see the first row in Table 11). The model is built with only five variables, Deviationfreq, Pitchmax, Pitchmaxt, Pitchd and Ppp selected by the wrapper approach. The missing arcs between predictor variables and the arcs from the class to the predictor variables encode the assumption of conditional independence underlying naive Bayes. Figure 4 also shows the parameters, $p(c)$ and the mean and standard deviation of the Gaussian distributions $f_j(x_j|c)$ in Eq. (1).

Table 10 Context prediction per dog

	Alone 43.40 %	Ball 48.85 %	Fight 76.34 %	Food 51.89 %	Play 49.44 %	Stranger 60.19 %	Walk 35.48 %
Dog1	–	–	–	–	76.00 %	68.00 %	–
Dog2	–	–	–	64.00 %	–	54.00 %	–
Dog3	52.00 %	44.00 %	56.00 %	–	–	60.00 %	–
Dog4	44.00 %	60.00 %	–	–	–	–	–
Dog5	–	–	98.00 %	–	–	70.00 %	–
Dog6	–	36.00 %	76.00 %	44.00 %	12.00 %	–	–
Dog7	52.94 %	29.41 %	52.94 %	17.65 %	–	35.29 %	41.18 %
Dog8	14.29 %	64.29 %	57.14 %	64.29 %	21.43 %	50.00 %	14.29 %

Accuracies of the best model in Table 9 for each of the eight dogs. The overall accuracy of this model over the eight dogs is 55.50 %

Table 11 Context discrimination: A model per dog

Dog	Model	Accuracy	Context
Dog5	Naive Bayes wrapper		
	k-Nearest neighbors wrapper	100.00 %	Fight · Stranger
Dog1	*k*-Nearest neighbors wrapper	97.00 %	Play · Stranger
Dog2	Logistic regression wrapper	86.00 %	Food · Stranger
Dog4	Naive Bayes wrapper	78.00 %	Alone · Ball
Dog3	*k*-Nearest neighbors wrapper	74.00 %	Alone · Ball · Fight · Stranger
Dog6	Logistic regression wrapper	73.00 %	Ball · Fight · Food · Play
Dog7	Naive Bayes wrapper	59.80 %	Alone · Ball · Fight · Food · Stranger · Walk
Dog8	*k*-Nearest neighbors wrapper	66.98 %	Alone · Ball · Fight · Food · Play · Stranger · Walk

Summary of the best models, accuracies and corresponding contexts for each dog. Dogs are organized by number of contexts and then by model accuracy

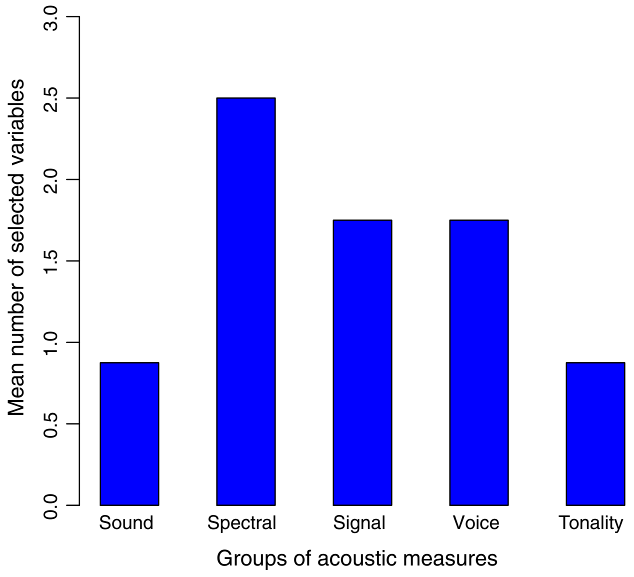

Fig. 3 Mean number of variables (*Y*-axis) selected by the best models per dog when predicting contexts (listed in Table 11), for each of the five groups of acoustic measures (*X*-axis): sound energy, spectral energy, source signal, voice cycles and tonality. Each of these groups of acoustic measures contain 6, 8, 9, 3 and 4 variables, respectively

Figure 5 displays the classification tree model which performed second best for dog1, with two observed contexts, Play and Stranger (see the second row in Table 11). Note that three variables are required: Energydiff, Harmmean and Ppj. Thus, if for a given bark, Energydiff = 10, Harmmean = 15 and Ppj = 0.05, then the dog is classified as barking at a stranger.

Figure 6 shows the 100 barks recorded for dog1, represented as a point in the 3-D space of three of the five variables selected by the best model, a *k*-nearest neighbors wrapper. Barks in the Play context are colored blue (dark), whereas Stranger is shaded red (light). A new bark (an asterisk in the figure) would be classified as the context of its nearest neighbor bark, i.e., Play in this 3-D space, although its nearest neighbor bark should be computed in the 5-D space, also including variables Deviationfreq and Harmmean.

Table 12 includes the details of the logistic regression model which performed best for dog2, with two observed contexts, Food and Stranger (see the second row in Table 11). This model is built from the five predictor variables in the first column. The regression coefficients $\beta_j^{(c)}$ for these variables would be used as in Eq. (2) to compute the posterior probability that yields the predicted class.

Individual

A single model for all contexts. Table 13 shows the results of a single model learned from the 800 barks for discriminating among the 8 dogs.

k-nearest neighbors wrapper is the best model, as in the three previous classification problems, with an extremely high accuracy, 67.63 %, in an 8 multi-class problem. Thus, feature subset selection methods have been proved to produce improvements in model performance.

The true positive rate for each of the classes can be computed from Table 14. Dogs numbers 8, 5 and 7 have high true positive rates: 0.77, 0.75 and 0.74, respectively. In contrast, dogs number 6 and 3 have the lowest true positive rates 0.51 and 0.58, respectively.

A model per context. More refined context-specific models are built here. By selecting bark sounds from the same context, the corresponding model will classify the individual dog for that context. Thus, a total number of 7 contexts (and their corresponding 12 × 7 models) have been considered, where the accuracy of the best model for each context is shown (see Table 15).

Note that the model accuracies for identifying dogs have increased to an 80–100 % range compared with the 67.63 % achieved by the global model learned from a

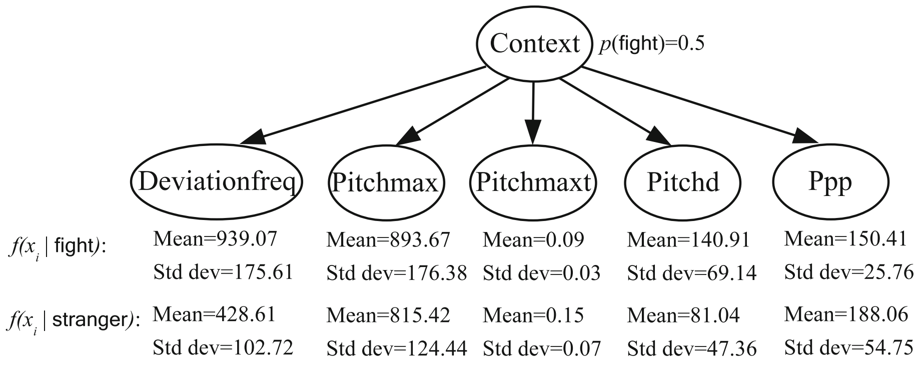

Fig. 4 Example of a naive Bayes wrapper model. It corresponds to the best model for context classification in dog5

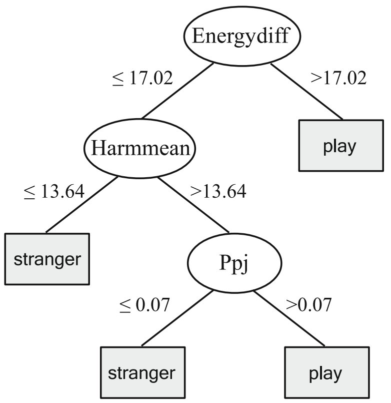

Fig. 5 Example of a classification tree wrapper model. It corresponds to the second best model for context classification in dog1

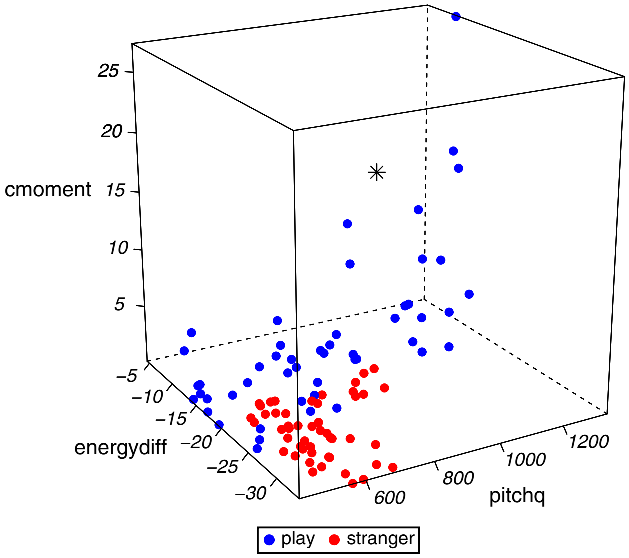

Fig. 6 Example of a k-nearest neighbors wrapper model. It corresponds to the best model for context classification in dog1 (Cmoment scale is divided by 10^9). Classification of a hypothetical bark (asterisk)

database with all the contexts. We now have fewer dogs to be identified, from 2 dogs for the Walk context to 5 dogs for Ball and Fight contexts, whereas the global model had the harder problem of identifying 8 dogs. Although the problem is easier because there are fewer class variable values, barking is expected to be homogeneous in a fixed context, which complicates correct dog identification.

Table 12 Example of parameter values of a logistic regression model

Variable X_j	$\beta_j^{(\text{Food})}$
Kurtosis	−0.0008
Pitchd	−0.0002
Pitchslope	0.0001
Ppp	−0.0143
Ppm	−7,424.9241
Intercept (β_0)	31.5997

It corresponds to the best model for context classification in dog2

Table 13 Individual prediction

	All	Filter	Wrapper
Naive Bayes	54.50 %	55.63 %	63.00 %
Classification tree	53.13 %	51.37 %	56.37 %
k-Nearest neighbors	63.87 %	58.62 %	**67.63 %**
Logistic regression	63.00 %	61.75 %	65.75 %

Accuracies of the twelve models: three selection feature methods for each of the four supervised classifiers

Table 14 Confusion matrix for the best model, k-nearest neighbors wrapper, identifying individual dogs

	Dog	Predicted class							
		1	2	3	4	5	6	7	8
Real class	1	68	10	8	5	0	5	3	1
	2	6	71	5	1	5	8	2	2
	3	9	8	58	6	2	14	1	2
	4	7	3	4	67	2	9	2	6
	5	1	6	3	2	75	7	6	0
	6	5	12	11	6	6	51	9	0
	7	2	3	2	5	5	10	74	1
	8	1	4	1	8	1	4	2	77

Table 15 Summary of the results of classifying individuals by context

Context	No. barks	No. dogs	Accuracy
Alone	106	4	94.34 %
Ball	131	5	80.92 %
Fight	131	5	88.55 %
Food	106	4	87.74 %
Play	89	3	97.75 %
Stranger	206	5	80.58 %
Walk	31	2	100.00 %

k-nearest neighbors performed best for Alone, Ball, Play and Walk contexts, naive Bayes for Fight and Walk, logistic regression for Stranger and Walk, and classification trees for Food context. All best models corresponded to a wrapper feature subset selection strategy

Predictor variables of sex, age, context and individual

The number of selected variables in the best models (see Table 16) represents about 50 % of the 29 initial variables. These numbers were 12 for sex, 15 for age, 16 for context and 18 for individual. It is remarkable that some variables, like Ltasm, Ltass, Pitchmint, Pitchslopenojump and Harmmax were never chosen. On the other hand, the following six variables occur in all four models: Energy, Ltasp, Deviationfreq, Skewness, Pitchq and Harmmean.

Harmdev appears to be specific for determining dog sex, since it was not selected in the rest of the problems. This also applies to Pitchd, only selected for discriminating dog age and to Pitchmaxt for individual determination.

Considering the blockwise organization of predictor variables in Table 3, sound energy (first block), source signal (third block) and tonality (fifth block) measurements are sparsely selected compared to a denser selection in the remaining blocks.

Table 16 Predictor variables of sex, age, context and individual classification problems from the best model, *k*-nearest neighbors wrapper

Var	Name	Sex	Age	Context	Individual
X_1	Energy	x	x	x	x
X_{10}	Loudness			x	x
X_{23}	Ltasm				
X_{24}	Ltass				
X_{25}	Ltasp	x	x	x	x
X_{26}	Ltasd	x	x		
X_2	Banddensity		x	x	x
X_3	Centerofgravityfreq	x		x	x
X_4	Deviationfreq	x	x	x	x
X_5	Skewness	x	x	x	x
X_6	Kurtosis		x	x	x
X_7	Cmoment		x	x	x
X_8	Energydiff		x	x	x
X_9	Densitydiff		x		
X_{11}	Pitchm	x			x
X_{12}	Pitchmin				x
X_{13}	Pitchmax			x	x
X_{14}	Pitchmint				
X_{15}	Pitchmaxt				x
X_{16}	Pitchd		x		
X_{17}	Pitchq	x	x	x	x
X_{18}	Pitchslope				x
X_{19}	Pitchslopenojump				
X_{20}	Ppp	x	x	x	
X_{21}	Ppm		x	x	x
X_{22}	Ppj	x		x	
X_{27}	Harmmax				
X_{28}	Harmmean	x	x	x	x
X_{29}	Harmdev	x			

The accuracies of these four models are 85.13 % for sex classification (Table 5), 80.25 % for age prediction (Table 7), 55.50 % for context categorization (Table 9) and 67.63 % for individual recognition (Table 13)

Discussion

This work has empirically demonstrated the usefulness of supervised classification machine learning methods for inferring some characteristics of dogs from the acoustic measurements given by their barks. From the four classification methods considered, *k*-nearest neighbors outperformed naive Bayes, classification trees and logistic regression. Also, the wrapper feature subset selection method provided significant improvements over a filter selection or no-selection (all variables are kept).

A solution for two prediction problems, sex and age, never previously considered in the literature has been presented. The best of the 12 resulting models in this study was able to predict dog sex in 85.13 % of the cases. The age of the dog, categorized as Young, Adult and Old, was inferred correctly in 80.25 % of the cases. An issue to be considered as future work is the prediction of age as a continuous variable, using a kind of regression task.

Determining the context of the dog bark, with seven possible situations, is a more difficult problem than classification by sex and age. However, it was successfully solved for 55.50 % of the bark cases. This is an improvement on the results presented in Molnár et al. (2008), where for six possible contexts the best model yielded a 43 % success rate. With an accuracy rate of 63 % for classifying three possible contexts, our results are similar to the findings reported by Yin and McCowan (2004). In addition, a model for each of the eight dogs with two or more different contexts was induced from the barks associated with this specific dog. Thus, a total of 12 × 8 models have been considered. For almost all dogs, the *k*-nearest neighbor model was the most successful, although naive Bayes, logistic regression and classification tree models provided the best accuracy results for some dogs. As a tendency, the wrapper feature subset selection strategy provided the best results. Model accuracy ranges from 59.80 to 100 %.

The individual identification, a hard classification problem with eight possible categories, produced up to 67.63 % accuracy in the best model. This result is extremely good when compared to the 52 % reported in Molnár et al. (2008) for 14 dogs, and the 40 % achieved by Yin and McCowan (2004) for a 10-dog problem. When the dog identification is performed within each context, the accuracies of the best models are in the interval [80.58 %, 100 %].

Recent ethological research on dog barking revealed several features of the most characteristic acoustic communication type of dogs which proved that barks serve as a complex source of information for listeners (Yin and McCowan 2004; Pongrácz et al. 2005, 2006). In experiments where human participants evaluated the pre-recorded dog barks, both the context and the possible inner state of the signaling animals were classified with substantial success rates. However, the role of dog barks in dog–dog communication remained (and still remains) somewhat obscure, as there is a shortage of convincing field data for the usage of barks during intraspecific communication of dogs, though see Pongrácz et al. (2014) for some positive evidence. The present study provides an alternative approach for discovering the potential information content encoded in dog barks. If one can prove that dog barks carry consistent cues encoding such features of the caller such as its sex, age or identity, this can prove indirectly that barks can serve as relevant sources of information to receivers that are able to decipher these types of information.

Previously, it was known that dogs can differentiate between individuals and contexts if they hear barks of other dogs in experiments based on the habituation–dishabituation paradigm (Maros et al. 2008; Molnár et al. 2009). Our new results provide some possible details of how such a capacity for recognition might work. If dogs are sensitive to the sex-, age- and identity-specific details of barks, this can serve as an acoustic basis for the cognitive task of discriminating between or recognition of individuals. Although in dogs sex-related information is mostly (thought to be) transferred via chemical compounds (Goodwin et al. 1979), theoretically it would be adaptive if a dog could survey the gender of the other dogs living nearby (or farther) on the basis of hearing their barks as well. Deciphering the age of an individual based on their vocalizations would be also beneficial in a highly social species, where age can be relevant in determining social rank, reproductive status or fighting potential (Mech 1999).

Recognition of the context of barks was the least successful task for our supervised learning methods. Although present methods exceeded the accuracy of both the previously employed machine learning approach (Molnár et al. 2008) and the adult human listeners' success rate (Pongrácz et al. 2005), this accuracy still lags behind the other variables analyzed in this study. It is also true that human listeners perform almost as successfully when recognizing the context as the computerized models. The reason behind this result may be that the individual variability of dog barks can be considerable especially in particular contexts (such as before the walk, or asking for a toy/food). Another reason for the relatively low success rate of context recognition may be that while the human listeners received short bark sequences, the computer worked with individual bark sounds. Therefore, the inter-bark interval served as an additional source of information for the humans (Pongrácz et al. 2005, 2006), while this parameter was not involved in the computerized analysis. For humans at least, the inter-bark interval also seemed to be an important source of information when discriminating between individual dogs, as their performance improved with the length of bark sequences they received (Molnár et al. 2006).

Supervised classification machine learning methods do not only provide indirect proof about the rich and biologically relevant information content of dog barks, but they also offer a promising tool for applied research, too. For example, evaluating dog behavior has great importance for various organizations, as well as professionals and dog enthusiasts. Recognizing unnecessarily aggressive dogs can be a challenge for the personnel of dog shelters as well as for correspondents of breed clubs and for the experts of legal bodies (Netto and Planta 1997; Serpell and Hsu 2001). Similarly, diagnosing particular behavioral abnormalities that can cause serious welfare issues for dogs, such as separation anxiety, can present a difficult task when the goal is to tell apart 'everyday' and chronic stress reactions in a dog (Overall et al. 2001). Behavioral evaluation usually does not cover the qualitative analysis of vocalizations in these cases. However, this could be addressed if a reliable and easy to use acoustic analytic software could serve as an aid for behavioral professionals. With such a method, following a rigorous validating protocol, acoustic features indicative of high levels of aggression, fear, distress, etc. could be recognized in the subjects' vocalizations.

The limitations of the supervised classification models presented in this paper concern the standard problems with the sample representativeness and the assumptions upon which the models rely. On the other hand, the generality of the four methods makes them directly applicable to other species. In addition, all the dogs in this study were of the same breed, so our classifiers do not take any advantage of the different patterns expected from the diversity of breeds.

An interesting problem for the near future would be to see whether these methods would work for other breeds or for a mixed breed group. Also, simultaneously classifying the four dog features, sex, age, context and individual, might be of interest. This issue falls into a category of a new problem type called multi-dimensional classification problems (Bielza et al. 2011; Borchani et al. 2012; Sucar et al. 2014), where the dependence between the four class variables is relevant.

Acknowledgments This work has been partially supported by the Spanish Ministry of Economy and Competitiveness, projects Cajal Blue Brain (C080020-09; the Spanish partner of the Blue Brain Project initiative from EPFL) and TIN2013-41592-P, by the János Bolyai Research Scholarship from the Hungarian Academy of Sciences, and co-financed by a grant from OTKA K82020. The authors are thankful to Celeste R. Pongrácz for the English proofreading of this manuscript and to Nikolett Czinege for help in preparing the bark databases.

References

Acevedo M, Corrada-Bravo C, Corrada-Bravo H, Villanueva-Rivera L, Aide T (2009) Automated classification of bird and amphibian calls using machine learning: a comparison of methods. Ecol Inform 4(4):206–214

Adachi I, Kuwahata H, Fujita K (2007) Dogs recall their owner's face upon hearing the owner's voice. Anim Cogn 10:17–21

Adams M, Law B, Gibson M (2010) Reliable automation of bat call identification for Eastern New South Wales, Australia, using classification trees and AnaScheme software. Acta Chiropterol 12(1):231–245

Armitage D, Ober H (2010) A comparison of supervised learning techniques in the classification of bat echolocation calls. Ecol Inform 5(6):465–473

Au W, Andersen L, Roitblat ARH, Nachtigall P (1995) Neural network modeling of a dolphin's sonar discrimination capabilities. J Acoust Soc Am 98:43–50

Bálint A, Faragó T, Dóka A, Miklósi A, Pongrácz P (2013) "Beware, I am big and non-dangerous!"—playfully growling dogs are perceived larger than their actual size by their canine audience. Appl Anim Behav Sci 148:128–137

Bielza C, Li G, Larrañaga P (2011) Multi-dimensional classification with Bayesian networks. Int J Approx Reason 52:705–727

Blumstein D, Munos O (2005) Individual, age and sex-specific information is contained in yellow-bellied marmot alarm calls. Anim Behav 69(2):353–361

Borchani H, Bielza C, Martínez-Martín P, Larrañaga P (2012) Markov blanket-based approach for learning multi-dimensional Bayesian network classifiers: an application to predict the European Quality of Life-5 Dimensions (EQ-5D) from the 39-item Parkinson's Disease Questionnaire (PDQ-39). J Biomed Inform 45:1175–1184

Britzke E, Duchamp J, Murray K, Swihart R, Robbins L (2011) Acoustic identification of bats in the Eastern United States: a comparison of parametric and nonparametric methods. J Wildl Manage 75(3):660–667

Charrier I, Aubin T, Mathevon N (2010) Mother-calf vocal communication in Atlantic walrus: a first field experimental study. Anim Cogn 13(3):471–482

Cheng J, Sun Y, Ji L (2010) A call-independent and automatic acoustic system for the individual recognition of animals: a novel model using four passerines. Pattern Recogn 43(11):3846–3852

Chesmore E (2001) Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. Appl Acoust 62(12):1359–1374

Clemins P (2005) Automatic Classification of Animal Vocalizations. PhD thesis, Marquete University

Cohen J, Fox M (1976) Vocalizations in wild canids and possible effects of domestication. Behav Process 1:77–92

Coppinger R, Feinstein M (1991) "Hark! Hark! the dogs bark*ldots*" and bark and hark. Smithonian 21:119–128

Druzhkova A, Thalmann O, Trifonov V, Leonard J, Vorobieva N, Ovodov N, ASGraphodatsky, Wayne R (2013) Ancient DNA analysis affirms the canid from Altai as a primitive dog. PLoS ONE 8(e57):754

Duda R, Hart P, Stork D (2001) Pattern classification. Wiley, New York

Fant G (1976) Acoustic theory of speech production. Mouton De Gruyter.

Faragó T, Pongrácz P, Miklósi A, Huber L, Virányi Z, Range F (2010a) Dogs' expectation about signalers' body size by virtue of their growls. PLoS ONE 5(12):e15,175

Faragó T, Pongrácz P, Range F, Virányi Z, Miklósi A (2010b) The bone is mine': affective and referential aspects of dog growls. Anim Behav 79(4):917–925

Feddersen-Petersen DU (2000) Vocalization of European wolves (Canis lupus lupus l.) and various dog breeds (Canis lupus f. fam.). Arch Tierz Dummerstorf 43(4):387–397

Fix E, Hodges JL (1951) Discriminatory analysis. Nonparametric discrimination: consistency properties. USAF Sch Aviat Med 4:261–279

Frommolt KH, Goltsman M, MacDonald D (2003) Barking foxes, Alopex lagopus: field experiments in individual recognition in a territorial mammal. Anim Behav 65:509–518

Goodwin M, Gooding KM, Regnier F (1979) Sex pheromone in the dog. Science 203:559–561

Gunasekaran S, Revathy K (2011) Automatic recognition and retrieval of wild animal vocalizations. Int J Comput Theor Eng 3(1):136–140

Hall M (1999) Correlation-based feature selection for machine learning. PhD thesis, Department of Computer Science, University of Waikato, UK

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I (2009) The WEKA data mining software: an update. SIGKDD Explor 11(1):10–18

Hartwig S (2005) Individual acoustic identification as a non-invasive conservation tool: an approach to the conservation of the African wild dog Lycaon pictus (Temminck, 1820). Bioacoustics 15:35–50

Hecht J, Miklósi A, Gácsi M (2012) Behavioral assessment and owner perceptions of behaviors associated with guilt in dogs. Appl Anim Behav Sci 139:134–142

Hunag C, Yang Y, Yang D, Chen Y (2009) Frog classification using machine learning techniques. Expert Syst Appl 36(2):3737–3743

Jain A, Murty MN, Flynn P (1999) Data clustering: a review. ACM Comput Surv 31(3):264–323

Le Cessie S, van Houwelingen JC (1992) Ridge estimators in logistic regression. Appl Stat 41(1):191–201

Liu H, Motoda H (1998) Feature selection for knowledge discovery and data mining. Kluwer Academic, Dordrecht

Lord K, Feinstein M, Coppinger R (2000) Barking and mobbing. Behav Process 81:358–368

Manser M, Seyfarth R, Cheney D (2002) Suricate alarm calls signal predator class and urgency. Trends Cogn Sci 6(2):55–57

Maros K, Pongrácz P, Bárdos G, Molnár C, Faragó T, Miklósi A (2008) Dogs can discriminate barks from different situations. Appl Anim Behav Sci 114:159–167

Mazzini F, Townsend SW, Virányi Z, Range F (2013) Wolf howling is mediated by relationship quality rather than underlying emotional stress. Curr Biol 23:1677–1680

McConnell PB (1990) Acoustic structure and receiver response in domestic dogs, Canis familiaris. Anim Behav 39:897–904

McConnell PB, Baylis JR (1985) Interspecific communication in cooperative herding: acoustic and visual signals from human shepherds and herding dogs. Z Tierpsychol 67:302–382

Mech LD (1999) Alpha status, dominance and division of labor in wolf packs. Can J Zool 77:1196–1203

Meints K, Racca A, Hickey N (2010) Child-dog misunderstandings: children misinterpret dogs' facial expressions. In: Proceedings of the 2nd Canine Science Forum, p 99

Miklósi A, Polgárdi R, Topál J, Csányi V (2000) Intentional behaviour in dog-human communication: an experimental analysis of "showing" behaviour in the dog. Anim Cogn 3:159–166

Minsky M (1961) Steps toward artificial intelligence. T Ins Radio Eng 49:8–30

Molnár C, Pongrácz P, Dóka A, Miklósi A (2006) Can humans discriminate between dogs on the base of the acoustic parameters of barks? Behav Process 73:76–83

Molnár C, Kaplan F, Roy P, Pachet F, Pongrácz P, Dóka A, Moklósi A (2008) Classification of dog barks: a machine learning approach. Anim Cogn 11:389–400

Molnár C, Pongrácz P, Faragó T, Dóka A, Miklósi A (2009) Dogs discriminate between barks: the effect of context and identity of the caller. Behav Process 82(2):198–201

Morton E (1977) On the occurrence and significance of motivation—structural rules in some bird and mammal sounds. Am Nat 111:855–869

Netto W, Planta D (1997) Behavioural testing for aggression in the domestic dog. Appl Anim Behav Sci 52:243–263

Overall K, Dunham A, Frank D (2001) Frequency of nonspecific clinical signs in dogs with separation anxiety, thunderstorm phobia, and noise phobia, alone or in combination. J Am Vet Med Assoc 219:467–473

Parsons S (2001) Identification of New Zeeland bats (Chalinobus tuberculatus and Mystacina tuberculata) in flight from analysis of echolocation calls by artificial neural networks. J Zool 253(4):447–456

Parsons S, Jones G (2000) Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. J Exp Biol 203(17):2641–2656

Pongrácz P, Molnár C, Miklósi A, Csányi V (2005) Human listeners are able to classify dog (canis familiaris) barks recorded in different situations. J Comp Psychol 119:136–144

Pongrácz P, Molnár C, Miklósi A (2006) Acoustic parameters of dog barks carry emotional information for humans. Appl Anim Behav Sci 100:228–240

Pongrácz P, Molnár C, Miklósi A (2010) Barking in family dogs: an ethological approach. Vet J 183:141–147

Pongrácz P, Szabó E, Kis A, Péter A, Miklósi A (2014) More than noise? Field investigations of intraspecific acoustic communication in dogs (Canis familiaris). Appl Anim Behav Sci (in press)

Quinlan R (1993) C4.5: programs for machine learning. Morgan Kaufmann

Reid P (2009) Adapting to the human world: dogs' responsiveness to our social cues. Behav Process 80:325–333

Roch M, Soldevilla M, Hoenigman R, Wiggins S, Hidebrand J (2008) Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes. Can Acoust 36(1):41–47

Root-Gutteridge H, Bencsik M, Chebli M, Gentle L, Terrell-Nield C, Bourit A, Yarnell RW (2013) Improving individual identification in captive Eastern grey wolves (Canis lupus lycaon) using the time course of howl amplitudes. Bioacoustics 23(1):39–53

Saeys Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23(19):2507–2517

Serpell J, Hsu Y (2001) Development and validation of a novel method for evaluating behavior and temperament in guide dogs. Appl Anim Behav Sci 72:347–364

Smith A, Birnie A, Lane K, French J (2009) Production and perception of sex differences in vocalizations of wied's black-tufted-ear marmosets (callithrix kuhlii). Am J Primatol 71:324–332

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J R Stat Soc B 36(2):111–147

Sucar E, Bielza C, Morales E, Hernandez-Leal P, Zaragoza J, Larrañaga P (2014) Multi-label classification with Bayesian network-based chain classifiers. Pattern Recogn Lett 41:14–22

Taylor A, Reby D, McComb K (2008) Human listeners attend to size information in domestic dog growls. J Acoust Soc Am 123(5):2903–2909

Taylor A, Reby D, McComb K (2009) Context-related variation in the vocal growling behaviour of the domestic dog (Canis familiaris). Ethology 115(10):905–915

Taylor A, Reby D, McComb K (2010) Size communication in domestic dog, Canis familiaris, growls. Anim Behav 79(1):205–210

Téglás E, Gergely A, Kupán K, Miklósi A, Topál J (2012) Dogs' gaze following is tuned to human communicative signals. Curr Biol 22:1–4

Tembrock G (1976) Canid vocalizations. Behav Process 1:57–75

Volodin I, Volodina E, Klenova A, Filatova O (2005) Individual and sexual differences in the calls of the monomorphic white-faced whistling duck dendrocygna viduata. Acta Ornithol 40:43–52

Wan M, Bolger N, Champagne F (2012) Human perception of fear in dogs varies according to experience with dogs. PLoS ONE 7(e51):775

Yeon SC (2007) The vocal communication of canines. J Vet Behav 2:141–144

Yin S (2002) A new perspective on barking in dogs (Canis familiaris). J Comp Psychol 116:189–193

Yin S, McCowan B (2004) Barking in domestic dogs: context specificity and individual identification. Anim Behav 68:343–355

Yovel Y, Au WWL (2010) How can dolphins recognize fish according to their echoes? A statistical analysis of fish echoes. PLoS ONE 5(11):e14,054