Anim Cogn
DOI 10.1007/s10071-007-0129-9

ORIGINAL PAPER

Classification of dog barks: a machine learning approach

Csaba Molnár · Frédéric Kaplan · Pierre Roy · François Pachet · Péter Pongrácz · Antal Dóka · Ádám Miklósi

Received: 6 July 2006 / Revised: 23 November 2007 / Accepted: 13 December 2007

Abstract In this study we analyzed the possible context-specific and individual-specific features of dog barks using a new machine-learning algorithm. A pool containing more than 6,000 barks, which were recorded in six different communicative situations was used as the sound sample. The algorithm's task was to learn which acoustic features of the barks, which were recorded in different contexts and from different individuals, could be distinguished from another. The program conducted this task by analyzing barks emitted in previously identified contexts by identified dogs. After the best feature set had been obtained (with which the highest identification rate was achieved), the efficiency of the algorithm was tested in a classification task in which unknown barks were analyzed. The recognition rates we found were highly above chance level: the algorithm could categorize the barks according to their recorded situation with an efficiency of 43% and with an efficiency of 52% of the barking individuals. These findings suggest that dog barks have context-specific and individual-specific acoustic features. In our opinion, this machine learning method may provide an efficient tool for analyzing acoustic data in various behavioral studies.

Keywords Acoustic communication · Dog barks · Machine learning · Genetic programming

We cite several conference proceedings because in computer science the conference proceedings are more important than in biology. For engineers, it is essential to show their colleagues that their product (e.g. software) is actually working, so in this field of science the main forums for scientific discussion are the conferences.

C. Molnár (✉) · P. Pongrácz · A. Dóka · Á. Miklósi
Department of Ethology, Eötvös Loránd University,
Pázmány Péter sétány 1/C, 1117 Budapest, Hungary
e-mail: molcsa@gmail.com

F. Kaplan
Ecole Polytechnique Fédérale de Lausanne,
CRAFT, CE 1 628 Station 1, 1015 Lausanne, Switzerland

P. Roy · F. Pachet
Sony Computer Science Laboratory,
Paris, 6 rue Amyot, 75005 Paris, France

Introduction

In this paper, we report the results of the first acoustic analysis and classification of companion dog barks using machine learning algorithms. Earlier we found that humans have the ability to categorize various barks and associate them with appropriate emotional content by merely listening to them (Pongrácz et al. 2005). Humans with different dog experience levels showed similar trends in categorization of the possible inner state of the given barking dog. In another study we have shown that human perception of the motivational state in dogs is influenced by acoustic parameters in the barks (Pongrácz et al. 2006). In contrast, humans showed only modest accuracy in discriminating between individual dogs by only hearing their barks (Molnár et al. 2006).

In behavioral research, especially when data collection (for example acoustic signal analysis) is automated, the size of the data set is often extremely large. A promising approach to handle the resulting information overload is to automate the process of knowledge extraction using data mining techniques, thereby extracting novel information and relationships between biological features (Fielding 1999; Hatzivassiloglou et al. 2001). Machine learning techniques permit the building of models for a given classification task. Such models take the form of a mathematical function that can assign a given class (or label) to an unknown example. The machine is first trained on a set of labeled examples and then tested on a second set for which it must predict the labels. During the training phase, parameters of the models are tuned automatically by the learning algorithm in order to obtain the best classification performances on the training set (Bergeron 2003). Such methods which are based on machine learning algorithms have been applied with success in other biological disciplines, particularly molecular biology (King and Sternberg 1990; Muggleton et al. 1992), drug design (King et al. 1992, 1993), neurology (Zhang et al. 2005a, b; Mitchell 2005, Bogacz and Brown 2003) and ecology (Stockwell 2006; Recknagel 2001; Obach et al. 2001; Schleiter et al. 2001).

There are only a few cases where machine-learning techniques have been used in behavioral research, for example when such methods were applied for classification of dolphin sonar and other vocalizations (Au 1994; Kremliovsky et al. 1998). Artificial neural networks emulate the parallel processing ability and pattern recognition capability of the brain and the use of such methods in dolphin echolocation and other fields of bioacoustics is worthy (Au et al. 1995). Another field of behavioral studies where artificial intelligence methods were used is image recognition from video recordings for behavioral analysis (Burghardt et al. 2004; Burghardt and Calic 2006; Calic et al. 2005). These studies report experiments utilizing methods that can track objects (e.g. animal faces) in video footages. In the future, the developed versions of these techniques might be useful to automate the coding of recorded behavior of focal animals.

In this study, we analyzed more than 6,000 barks recorded in different situations from several individuals using a computerized method. The traditional computerized approach addressing audio classification problems is to combine the so called low level descriptors such as the one provided by the Mpeg7 standardization process to create acoustic *descriptors* (Monceaux et al. 2005; Cabral et al. 2005). The most relevant descriptors are then fed into machine-learning algorithms to produce *classifiers* (or *extractors*) whose performance is checked against perceptive tests. A new method we used in this study combines the construction of a new feature space and the search among several state-of-the-art machine-learning techniques (Pachet and Zils 2004). It is currently used in several projects within Sony Corporation and in collaboration with other research institutes. It has been proven to outperform other methods on problems related to song classification and recognition of urban noises (Defréville et al. 2006). The Weka system was used to search in the space of classification method. Weka implements a large number of performant machine learning schemes and facilitated comparisons between them (Witten and Eibe 1999).

In case of several vocal signals, the receivers can simultaneously gain information about the caller's identity, motivational state and communicative context (e. g. Gerhardt 1992; Hauser 1996). Studies have demonstrated that in various species there are characteristic and stable differences between the vocalizations of the individuals (e. g. spotted hyena, *Crocuta crocuta*: Holekamp et al. 1999; Arctic fox, *Alopex lagopus*: Frommolt et al. 2003; and domestic dog, *Canis familiaris*: Yin and McCowan 2004). Other studies have shown that certain vocal signals have context specific acoustic features (in meerkats, *Suricata suricatta*: Manser et al. 2002; Gunnison's prairie dogs, *Cynomys gunnisonii*; Slobodchikoff et al. 1991). In the past, only a few studies focused on the acoustic behavior of dogs (Tembrock 1976; Cohen and Fox 1976; Yin 2002) and none of these used methods based on artificial intelligence.

Methods

Subjects

Barks of the *Mudi* breed (a Hungarian sheepdog listed at the 238th Standard of the FCI (Fédération Cynologique International)) were used for this study. We recorded barks from 14 individuals, sex ratio (male/female): 4/10, age: 4.21 ± 3.17 years. The total sample size of barks analyzed was $N = 6{,}646$.

Source and collection of sound recordings

We collected bark recordings in seven different behavioral contexts, most of which could be arranged at the homes of the owners, with the exceptions of the "Fight" situation, which was staged at dog training schools, and the 'Alone' situation, which was staged on a street or in a park. The seven situations are as follows:

"Stranger" ($N = 1802$): The experimenter (male, age 23 years) was the stranger for all the dogs, and appeared in the garden of the owner or at the front door of his/her apartment in the absence of the owner. The experimenter recorded the barking of the dog during his appearance for 2–3 minutes.

"Fight" ($N = 1118$): For dogs to perform in this situation, the trainer encourages the dog to bark aggressively and to bite the glove on the trainer's arm. Meanwhile the owner keeps the dog on leash.

"Walk" ($N = 1231$): The owner was asked to behave as if he/she was preparing to go for a walk with the dog. For example, the owner took the leash of the dog in her/his hand and told the dog „We are leaving now".

"Alone" ($N = 752$): The owner tied the dog to a tree with a leash in a park and walked away, out of sight of the dog.

"Ball" ($N = 1001$): The owner held a ball (or some favorite toy of the dog) at a height of approximately 1.5 m in front of the dog.

"Play" ($N = 742$): The owner was asked to play with the dog a usual game, such as tug-of-war, chasing or wrestling. The experimenter recorded the barks emitted during this interaction.

For spectrograms of barks recorded in the above situations see Fig. 1.

Recording and preparing the sound material

Recordings were made with 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 within a 4–5 m distance from the dog. The experimenter tried to stand in front of the dog if it was possible. The recorded material was transferred to a computer, where it was digitalized with a 16-bit quantization and 44.10 kHz sampling rate using a TerraTec DMX 6fire 24/96 sound card.

As each recording possibly contained up to three or four barks, individual bark sounds were manually segmented and extracted. This process resulted in a final collection of 6,646 sound files containing only a single bark sound. As a consequence, this preparation excludes interval silences between two barks from the analysis, which have been shown to be potentially meaningful for humans (Pongrácz et al. 2006). As in any audio classification problem, it is clear that this segmentation phase is crucial for the interpretation of the result. This type of segmentation allows the ability to remove part of the background noise in the classification, which is an important issue for the reliability of the results presented. However, in the present study we did not test how the results would change with other segmentation strategies. Table 1 summarizes the structure of the database. The number of samples available for each dog and each situation varies in a significant manner.

Analyzing process

Phase 1: generation of a large number of descriptors adapted to a specific classification problem using EDS

EDS is an audio signal processing system that produces descriptors adapted to a particular audio classification problem. EDS contains an extensive library of basic operators (acoustic features) that it combines according to an evolutionary algorithm to produce descriptors that are much more specific to the problem we want to solve (for a selected list of operators see the Appendix). The way EDS combines basic operators is by mathematical composition of functions. EDS explores the space of all functions that can be expressed as a composition of basic operators. Such a function space is potentially huge, thus, searching for "interesting" functions is a highly combinatorial process.

EDS implements various artificial intelligence techniques to cope with this issue. First, a type system prevents the creation of malformed descriptors. Second, a library of weighted heuristics for operator composition is used to guide the elaboration of new descriptors. Finally, the descriptor creation is based on a genetic algorithm: EDS iteratively creates generations of descriptors; only the most efficient descriptors, according to their fitness, of one generation are kept as seeds for the next one, i.e. they are combined together according to "genetic mutations" (Koza 1992) to create new descriptors. The fitness used to assess descriptor's performance is based on class membership. For a given descriptor, a nearest-neighbor classifier is created that is trained and tested on non-overlapping datasets (33% of the total dataset is randomly selected to act as a training set, the rest is used as a testing set on the problem of context classification). The fitness of the descriptor is the performance of the associated classifier. This process goes on until very efficient descriptors have been devised by the

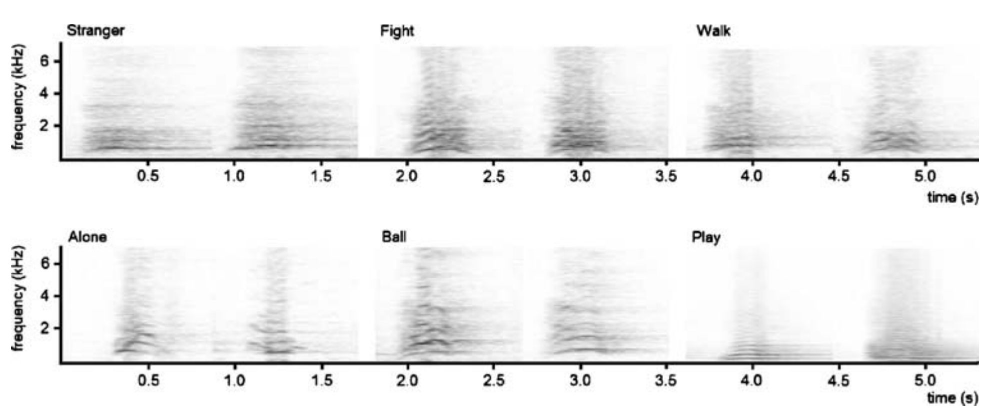

Fig. 1 Spectrograms of barks recorded in different contexts. The figure shows two barks recorded from different dogs for each context

Table 1 The number of barks collected from 14 dogs in 6 situations

	classes	Play	Fight	Alone	Stranger	Walk	Ball
dogs	**6646**	**742**	**1118**	**752**	**1802**	**1231**	**1001**
d24	**1524**	310	232	275	208	334	165
d12	**906**		335	142	184	62	183
d14	**766**		24	158	234	203	147
d23	**733**		129	130	140	143	191
d18	**627**	104		47	142	279	55
d09	**528**				471	19	38
d20	**521**	143	143			77	158
d16	**290**	33	82		107	59	9
d05	**261**	148			113		
d27	**128**		64		26	15	23
d26	**111**		28		59	24	
d25	**91**		11		48		32
d10	**85**				69	16	
d08	**75**	4	70		1		

system. Eventually, the best descriptors are fed into a machine-learning algorithm to produce classifiers (for further details of EDS analyzing process see Pachet and Zils 2004). For this study, we started with a set of basic signal processing operators and explored their combinations using EDS. A large number of descriptors adapted to the classification problem (situation recognition or dog recognition) were generated this way.

Phase 2: creation of an optimal subset of descriptors

We then searched for the subset of features that would most likely predict the class best. This involves a way to evaluate a feature set and a way to search the space of feature sets.

The motivation behind such reduction of the feature space is that many machine-learning techniques degrade in performance when presented with many features that are not necessary for predicting the output. This is especially true for Naive Bayes classifiers such as the one we used in this study, which assumes independence of feature and therefore suffers from correlated features. Two approaches are commonly used. In the "filter" approach, a feature subset is selected based on the properties of the data itself and is independent of the chosen classifier algorithm. In the "wrapper" approach, the classifier algorithm is used during the search of the feature subset (Kohavi and Sommerfield 1995).

The method used in this study belongs to the "wrapper" approach. To evaluate the feature subset, we used a simple Bayesian method trained over a part of a data set (roughly 2/3) and tested over another (1/3). For the situation classification problem, the test set did not contain any samples from dogs present in the train set. Respectively, for the dog classification problem, the train set did not contain any samples with situations present in the classification set. This prevents possible over fitting of the classification method to particular individuals, contexts or recording conditions. The choice of the Bayesian classifier was motivated by its good compromise between efficiency and simplicity and by the fact that it is mathematically well defined. The classifier used in this study is the NaiveBayesSimple classifier implemented in the Weka machine-learning library (Witten and Eibe 1999). The NaiveBayesSimple module fits a Gaussian (normal) distribution to each dimension, and then combines them by treating them as independent. This is a simple and classical method in statistical modelling, with clear semantics. However, it is possible that other classification methods outperform these classifiers for the two tasks considered.

The search in the space of the feature sets was conducted using Weka's "GreedyStepwise" search method (Witten and Eibe 1999). The search stopped when the addition or deletion of any remaining attributes resulted in a decrease in evaluation. With this method, for both classification problems considered, we reduced the space to a small number of very relevant features. During this phase, each classifier was trained over a part of a data set and tested over another for evaluating rapidly the quality of the attribute space. For the situation classification problem, the test set did not contain any samples for dogs present in the train set.

Respectively, for the dog classification problem, the train set did not contain any samples with situations present in the train set. This prevents from possible over fitting of the classification method to particular individuals, contexts or recording conditions.

Phase 3: Complete evaluation of recognition performance

Once an adapted attribute set was created for each classification problem, a complete evaluation of the classification performance with the Bayesian classifier was conducted. For the context recognition problem, 14 train/test sets were created, each of them corresponding to a test on the data of a single individual and training on all others. Similarly, for the individual recognition problem, 6 train/test sets were created, each of them corresponding to a test on a particular context and training on all others. Results were then aggregated and compared. Once again, this method prevents biases linked with particular recording conditions.

Results

Experiment 1: categorization of barks into contexts

In this experiment, we used the categorization algorithm for classification of barks into contexts. To construct the attribute space, a training set was constituted with recordings from nine dogs (d05, d08, d09, d10, d12, d14, d16, d18, d20—4,059 samples in total, 61% of the total number of samples) and a test set with recordings from five dogs (d23, d24, d25, d26, d27—2,587 samples in total, 39% of the total number of samples). Using the EDS algorithm and the optimal feature subset search method described above, the system converged on a set of seven features described in Table 2.

Table 3 presents the results of each independent individual test set (corresponding to the training of the classifiers to recognize situations with all the dogs except one and a test on the recordings of the remaining dog), as well as the overall performance (global and per context) and overall confusion matrix (for a simplified confusion matrix, see Table 4). The overall recognition rate is 43% (2,835 correctly classified instances over 6,646). As shown in Table 3, a random classification algorithm would have guessed only 18% (1,223 correctly classified instances over 6,646). The overall recognition rate was significantly higher than the recognition rate using a random algorithm (one-sample *t*-test: $t_{13} = 6.10$, $P < 0.001$) In order to correct for agreement that occurs by chance, we can calculate the kappa statistic by deducting the number of corrected predictions obtained by the random algorithm from the ones obtained by the Bayesian classifier (2,835 − 1,223 = 1,612) and comparing it with the possible total of 6,646 − 1,223 = 5,423 recognized instances. This results in an overall kappa of 30%.

The best recognition rates were achieved for the barks recorded in the "Fight" (74%, kappa = 68%) and "Stranger" (63%, kappa = 49%) contexts and poorest rate was achieved when categorizing the "Play" barks (6%,

Table 2 Constructed feature set optimal for context recognition

Name	Explanation
Spectralrolloff (derivation(x))	Rolloff: frequency below which 85% of the magnitude distribution of the spectrum is concentrated. Applied here to the first derivative of the signal.
Spectralflatness (square(x))	Measures the deviation of the spectrum of the signal from that of a flat spectrum. Here, it is applied to the signal to the square
Sqrt(RHF(derivation (sbs(derivation(x)))))	Computes the second derivative of the signal, and measures the High Frequency Ratio (ratio between high and low frequencies in the spectrum of the signal)
Rms(spectralskewness (split(x_512.0)))	Segments the signal every 512 sample, and measures the spectral skewness of each segment. Rms is the square root of the mean of the square values
Abs(max(spectralflatness (split(x_256.0))))	The absolute value of the max value of the Spectral Flatness (Measures the deviation of the spectrum of the signal from that of a flat spectrum) over 256-sample segments
Mean (formant (1,x))	Mean of first formant. The first formant is the first peak in the frequency spectrum (the lowest).
Deviation (harmonicity (x))	Standard deviation of harmonicity. The Harmonicity is the degree of acoustic periodicity, also called Harmonics-to-Noise Ratio (HNR) (measured in dB). The Harmonicity can be used to measure the signal-to-noise ratio of anything that generated a period signal. In speech analysis, it is also used to measure voice quality.

Table 3 Complete results of the situations classification task including overall recognition rate, results per situations, confusion matrix, confusion matrix for a random algorithm, kappa statistics, and details of individual test sets

	Total			Play			Fight			Alone			Stranger			Walk			Ball		
	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc
	2835	6646	43	41	742	6	824	1118	74	110	752	15	1131	1802	63	365	1231	30	364	1001	36
Dog test sets																					
d08	67	75	89	0	4	0	67	70	96				0	1	0						
d10	67	85	79										67	69	97	0	16	0			
d16	196	290	68	2	33	6	75	82	91				90	107	84	29	59	49	0	9	0
d09	355	528	67										320	471	68	10	19	53	25	38	66
d25	48	91	53				6	11	55				25	48	52				17	32	53
d26	58	111	52				23	28	82				35	59	59	0	24	0			
d12	439	906	48				239	335	71	5	142	4	114	184	62	3	62	5	78	183	43
d23	297	733	41				80	129	62	2	130	2	122	140	87	5	143	3	88	191	46
d27	51	128	40				34	64	53				9	26	35	2	15	13	6	23	26
d24	603	1,524	40	12	310	4	217	232	94	95	275	35	47	208	23	185	334	55	47	165	28
d14	266	766	35				0	24	0	4	158	3	159	234	68	62	203	31	41	147	28
d05	86	261	33	4	148	3							82	113	73						
d18	170	627	27	22	104	21				4	47	9	61	142	43	63	279	23	20	53	36
d20	132	521	25	1	143	1	83	143	58							6	77	8	42	158	27
	2,835	6,646	43																		
Confusion matrix																					
Play	41	742	6	**41**	**742**	**6**	93	742	13	29	742	4	186	742	25	294	742	40	99	742	13
Fight	824	1,118	74	8	1118	1	**824**	**1,118**	**74**	2	1,118	0	151	1,118	14	66	1,118	6	67	1,118	6
Alone	110	752	15	30	752	4	50	752	7	**110**	**752**	**15**	321	752	43	120	752	16	121	752	16
Stranger	1,131	1,802	63	13	1,802	1	233	1,802	13	72	1,802	4	**1,131**	**1,802**	**63**	112	1,802	6	241	1,802	13
Walk	365	1,231	30	139	1,231	11	136	1,231	11	66	1,231	5	373	1,231	30	**365**	**1,231**	**30**	152	1,231	12
Ball	364	1,001	36	20	1,001	2	73	1,001	7	50	1,001	5	378	1,001	38	116	1,001	12	**364**	**1,001**	**36**
	1,223	6,646	18																		
Random algorithm																					
Play	83	742	11	**83**	**742**	**11**	125	742	17	84	742	11	201	742	27	137	742	19	112	742	15
Fight	188	1,118	17	125	1,118	11	**188**	**1,118**	**17**	127	1118	11	303	1,118	27	207	1,118	19	168	1,118	15
Alone	85	752	11	84	752	11	127	752	17	**85**	**752**	**11**	204	752	27	139	752	19	113	752	15
Stranger	489	1,802	27	201	1,802	11	303	1,802	17	204	1,802	11	**489**	**1,802**	**27**	334	1,802	19	271	1,802	15
Walk	228	1,231	19	137	1,231	11	207	1,231	17	139	1,231	11	334	1,231	27	**228**	**1,231**	**19**	185	1,231	15
Ball	151	1,001	15	112	1,001	11	168	1,001	17	113	1,001	11	271	1,001	27	185	1,001	19%	**151**	**1,001**	**15**
Kappa (%)	30			**−6**			**68**			**4**			**49**			**14**			**25**		

The bold values indicate the percentages of correctly categorized contexts

Table 4 Simplified confusion matrix for the context categorization task

	play	fight	alone	stranger	walk	ball
play	6%	13%	4%	25%	40%	13%
fight	1%	74%	0%	14%	6%	6%
alone	4%	7%	15%	43%	16%	16%
stranger	1%	13%	4%	63%	6%	13%
walk	11%	11%	5%	30%	30%	12%
ball	2%	7%	5%	38%	12%	36%

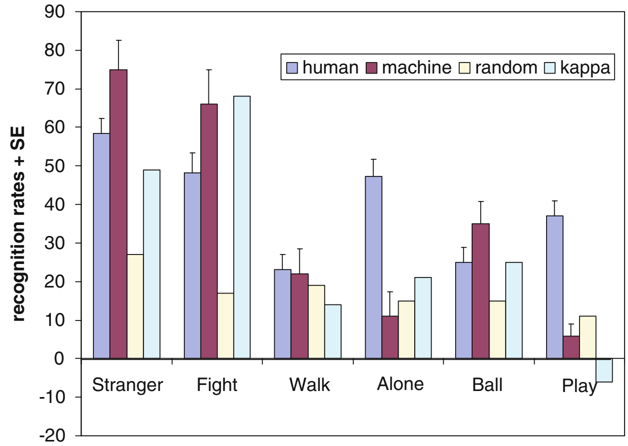

Fig. 2 Comparison of recognition rates obtained by the algorithm and the recognition rates (percentages of barks categorized correctly) of human listeners in one of our previous studies (Pongrácz et al. 2005). The level of chance is at 16.67%. *White* (*yellow*) *columns* represent the recognition rates obtained by a random algorithm and *light gray* (*light blue*) *columns* indicate the results of kappa statistics

kappa = −6%). We compared the recognition rates for barks in different contexts to each other and to the recognition level of the random algorithm as well. We found that the success rate of the algorithm was significantly higher in cases of "Stranger" (one-sample *t*-test: $t_{12} = 4.08$; $P < 0.01$), "Fight" ($t_9 = 5.45$; $P < 0.001$), and "Ball" ($t_9 = 3.55$; $P < 0.01$) situations, and we found no significant difference in cases of "Walk" ($t_{10} = 0.43$; $P = 0.67$), "Alone" ($t_4 = 0.06$; $P = 0.95$), and "Play" ($t_5 = 1.63$; $P = 0.16$) contexts. We compared the recognition rates of the algorithm for barks in different contexts with each other. We found that the algorithm was significantly more successful in categorizing "Stranger" and "Fight" barks than the others (one-way ANOVA: $F_{4,49} = 10.20$; $P < 0.001$) For recognition rates in all the situations and confusion matrix see Fig. 2 and Table -3. The confusion matrix shows the proportions of the correctly and incorrectly categorized barks in each situation. In the case of barks recorded in "Play" context, a relatively high proportion of erroneously classified barks were considered as "Walk" barks (40%). "Alone" barks are also confused with "Stranger" barks (43%). Important differences in the recognition rate can be observed in the individual dog test sets (maximum for d08, 89% and minimum for d20, 25%)

Experiment 2: recognition of dogs

Symmetrically with the previous experiment, for constructing an attribute space adapted to the recognition of individual dogs, we first separated the whole dataset into a training set with recordings corresponding to "Ball", "Stranger" and "Fight" barks (3,921 recordings, 59% of the whole dataset) and a test set corresponding to the "Alone", "Play" and "Walk" barks (2,725 recordings, 41% of the whole dataset) and vice versa. With such separation, dogs will have to be recognized in situations that the classifiers have not encountered before. Using the EDS algorithm and optimal feature subset search method described above, the system converged on a set of eight features described in Table 5.

Table 6 presents the results of each independent situation test set (classifiers trained to recognize dogs on all situations but one and tested on the remaining one), as well as the overall performance (global and per context) and global confusion matrix. For this task, the overall recognition rate is 52% (3,463 correctly classified instances over 6,646, kappa = 45%). We compared the recognition rates of the 14 individuals to the chance level (100/14 = 7.14), and found that the success rate of the algorithm was significantly higher than chance level (one-sample *t*-test: $t_{12} = 3.21$; $P < 0.01$). Important differences exist between dogs. Some dogs were easily recognized, as with d23 (733 recordings, kappa = 75%) or d24 (1,524 recordings, kappa = 69%), others were very poorly recognized, possibly because of the comparatively small number of recordings available as with d08 (75 recordings, kappa = −1%) or d10 (85 recordings, kappa = −1%). We should note that in the analyzed set of barks not every individual was represented by the same amount of collected barks. This can explain some of the differences of recognition rates between individual dogs (classes with too few training examples are poorly recognized).

Looking more precisely at the comparative performance of the different situation test sets, we compared the recogni-

Table 5 Constructed feature set optimal for dog recognition

Name	Explanation
Mean (formant (2,x))	Mean of the second formant. The second formant is the second peak in the frequency spectrum.
Mean (formant (4,x))	Mean of the fourth formant. The fourth formant is the fourth peak in the frequency spectrum.
Deviation (formant (3,x))	Standard deviation of the third formant. The third formant is the third peak in the frequency spectrum.
Abs(energy(0,500,x) − energy(500,4000,x))	Energy difference between two bands (0–500 Hz, 500–4,000 Hz).
Max(arcsinus(x))	Max value of the arcsinus of the signal
Mean(Zcr(split(x_512.0)))	Counts the zero-crossing for every segment of 512 samples, and returns the mean value
Mean (harmonicity (x))	Mean of Harmonicity. The Harmonicity is the degree of acoustic periodicity, also called Harmonics-to-Noise Ratio (HNR) (measured in dB). The Harmonicity can be used to measure the signal-to-noise ratio of anything that generated a period signal. In speech analysis, it is also used to measure voice quality.
Slope_no_jump (pitch(x))	Slope of the pitch without octave jump.

Table 6 Complete results of the dog classification task including overall recognition rate, results per dog, kappa statistics, and details of individual test sets

	Total			d05			d08			d09			d10			d12			d14			d16		
	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc
	3463	6646	52	10	261	4	0	75	0	131	528	25	0	85	0	531	906	59	259	766	34	86	290	30
Ball	650	1001	65							20	38	53				97	183	53	88	147	60	8	9	89
Alone	483	752	64								0					24	142	17	61	158	39			
Walk	786	1231	64							12	19	63	0	16	0	32	62	52	59	203	29	26	59	44
Play	448	742	60	4	148	3	0	4	0		0											27	33	82
Fight	553	1118	49				0	70	0		0					236	335	70	0	24	0	3	82	4
Stranger	543	1802	30	6	113	5	0	1	0	99	471	21	0	69	0	142	184	77	51	234	22	22	107	21
Kappa (%)	45			0			−1			18			−1			52			25			26		

	d18			d20			d23			d24			d25			d26			d27		
	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc	Cor	Tot	Perc
	403	627	64	269	521	52	573	733	78	1161	1524	76	1	91	1	8	111	7	31	128	24
Ball	33	55	60	96	158	61	175	191	92	129	165	78	1	32	3		0		3	23	13
Alone	27	47	57				123	130	95	248	275	90									
Walk	227	279	81	45	77	58	123	143	86	256	334	77				4	24	17	2	15	13
Play	100	104	96	57	143	40				260	310	84									
Fight				71	143	50	38	129	29	183	232	79	0	11	0	0	28	0	22	64	34
Stranger	16	142	11				114	140	81	85	208	41	0	48	0	4	59	7	4	26	15
Kappa (%)	61			48			75			69			0			6			23		

tion rates of the algorithm for barks recorded in different situations. We did not find significant differences among them (ANOVA: $F_{5,49} = 2.37$; $P = 0.056$), however it seems that recognizing a dog in the "Ball" (65%), "Alone" (64%), "Walk" (64%) and "Play" (60%) is easier than in the "Fight" (49%) and "Stranger" (30%). This is unlikely to be an effect of the particular separation of the train and test set used during the construction of the attribute space ("Ball", "Stranger", "Fight" for training, "Alone", "Play", "Walk" for testing) but we cannot discard this effect completely. This suggests the "Fight" and "Stranger" situations are indeed special contexts, both easier to recognize as context type but more difficult than other contexts for the recognition of individual dogs.

Discussion

Context classification

We have found that barks can be categorized into situations and different individuals' barks are distinguishable using a computerized method. The relative efficiencies of the software were comparable to human performance in the situations of "Stranger" and "Fight" (these were also the easiest for humans to discriminate, Pongrácz et al. 2005). In the situations of "Walk" and "Ball", the program outperformed humans, but in the "Play" and "Alone" contexts human showed better performance. The algorithm could categorize the barks into the correct contexts in 43% of the barks on average (kappa = 30%).

We have found earlier that humans could categorize the "Stranger", "Fight", and "Alone" barks with a relatively high accuracy rate (Pongrácz et al. 2005). Whereas in the "Play", "Walk" and "Ball" situations, their performance was less efficient, the "Stranger", "Fight" and "Alone" barks were categorized with a success rate of 45–50% while "Walk", "Ball" and "Play" barks were successfully categorized in 30% of cases. We concluded the reason for these findings might be that in the latter contexts the barks of different individuals are less similar to each other because there was not a strong selection affecting their acoustic features or that during the vocalization the motivational states of dogs in these situations varies very rapidly. In contrast, the "Stranger" and "Fight" barks, which could be dominated by agonistic tendencies, and the "Alone" (perhaps fearful) barks are more uniform among different individuals.

These findings suggest that barks could have context-specific features as it has been found in several other social species (e. g. meerkats: Manser et al. 2002; vervet monkeys, Seyfarth et al. 1980). This contradicts earlier arguments that during domestication, barking had lost the role in communicating distinct motivational states and/or contextual information as well as the callers' identity (e.g. Tembrock 1976; Cohen and Fox 1976). Yin (2004) was the first to point out that in some contexts barks are acoustically different which means that the acoustic features of the bark depend either on the motivational/emotional state and/or on the actual context. The achievements of the program clearly show that barks have the acoustic potential for being context-specific calls.

The different motivational states of dogs in aggressive, friendly or submissive contexts could result in acoustically different barks. Morton (1977) showed that several mammalian and bird species emit harsh, low frequency sounds in aggressive situations, and, in contrast, tonal, high frequency ones in submissive or friendly contexts. But these differences occur among qualitatively different vocalization types. Here in barks, dogs could modify their sounds within the same vocalization type, which suggest that barks are very easy to modify, so they could be an effective tool to communicate inner states in a flexible way. From this point of view, the dog's barks are different from wolf's barks since wolves bark only in very distinct contexts (e. g. in aggressive situations when an intruder attacks the den, see Feddersen-Petersen, 2000, Schassburger, 1993) These modifications could have happened as a result of evolutionary changes in the dog's vocalization system but another option is that dogs vary their barks according to their learned experiences.

Individual classification

The other important result was that the software was able to categorize the barking individuals with an efficiency of 52% (kappa = 50%). We found that the software categorized individuals at a higher level of success if they were barking in "Walk", "Alone", "Ball" and "Play" contexts, but the difference was not significant. This was very surprising because there was no previous evidence that barks may contain individual-specific information. Actually, applying a very similar task we have found that humans are not able to discriminate reliably between barks of different individuals (Molnár et al. 2006). Molnár et al. (2006) exposed subjects to two bark samples of Mudis recorded in the same context and their task was to guess if those two barks were recorded from the same or different dogs. Their performances did not pass the threshold of reliable discrimination.

Many authors hypothesized (Fitch et al. 2002) that the anatomical individual variability of the supralaryngeal vocal tract could be the primary source of cues used for individual recognition (Owren and Rendall 2003, Fitch et al. 2002). The reasoning of this argument is that the formant characteristics are better detectable if the call is more tonal which means that the power is concentrated in the harmonic pattern. Unpublished data collected by us suggest that barks in the situations "Alone", "Ball" and "Play" (where the algorithm was more successful) are more tonal.

In an individual-recognition task, the algorithm could recognize the dog with a higher efficiency when it was barking in "Ball", "Alone", "Walk" and "Play" situations in contrast to "Stranger" and "Fight" contexts where the success rate was lower. In a context-categorization task, the algorithm was most successful in categorizing "Stranger" and "Fight" barks and was relatively less successful in categorizing "Alone", "Play" and "Walk" barks. These two findings suggest that "Stranger" and "Fight" barks are less individual-specific and barks in the other situations are more distinguishable among individuals.

Comparisons of the performances of the humans and computerized algorithm could be useful to understand the effects in the background of human performances when categorizing the barks, although their performances can be compared at only a limited level due to methodological differences between the two experiments. Because of these differences unfortunately we cannot compare directly the performances of the algorithm and humans using statistical analyses. One important difference between this computerized analysis and a human perceptual test is that humans were exposed to several barks in sequence coming from the same context but the software used only "one" isolated bark to make its decision. Since in a previous study we found that the intervals between the barks have an effect on the decisions of listeners about the dog's motivational state (the barks with shorter intervals were considered more "aggressive" while barks with longer intervals were considered more "fearful", "desperate" and "happy"), one could presume that this acoustic feature might be different among various individuals' barks and in different contexts. In cases of "Alone" and "Play" barks the humans significantly outperformed the computer. A reason for this performance could be that in these situations the human listeners use the intervals between barks more extensively while the computer did not have the opportunity to use this feature for categorization.

Another difference between the human and computer tests is that the computerized algorithm was exposed to massive training on dog barks in the first phase of the experiment. Whereas humans could rely only on their previous experience of hearing dog barks and they did not have the opportunity to listen to training sound samples. The prior training could increase the success rate of listeners in case of barks recorded in "Ball", "Play" and "Walk" contexts because these contexts seemed to be less uniform. A future experiment should clarify whether the human performance levels can be enhanced by specific training on dog barks. According to our findings the computer significantly outperformed the humans in individual categorization task while in situation classifying tasks its performance was at a similar level to that of humans'. This suggests that prior training could have a significant effect on individual categorization abilities.

The main difference between performances of the computer algorithm and humans was that the software could reliably discriminate among individuals while humans could not (Molnár et al. 2006). It might mean that there are individual differences in barks of dogs but humans are not able to recognize them easily. In another study, we found that dogs can differentiate barks of different individuals. From the findings of these three studies, we hypothesized that there are individually distinctive features of barks but these characteristics are recognizable only by conspecifics and not by humans. However, that study was designed according to the "habituation-dishabituation paradigm", so the task of dogs was simpler than categorizing the barks.

In our opinion, the performance of the computer must have been based on distinct context-specific and individual-specific acoustic features of the barks. If these features emitted in a given situation had varied randomly they would not have been recognizable above chance level either for humans or a computer algorithm. This raises the question of what kind of information could be encoded in barks. During early domestication of dogs people might have preferred more vigilant dogs, which could alarm them when a stranger approached, defending the camp against intruders. If the dogs could recognize the barks of others, which were emitted in certain situations, it might improve the success and reliability of alerting. Hence there might have been a strong positive selection for dogs, which barked frequently especially if they could distinguish among barks of others.

It should be noted that the features we propose in this paper are the result of a particular experiment, with a particular database of dog bark sounds. The features found using the EDS system are better than the "conventional" features used in the audio classification literature. However, these features are not easily interpretable. Also, we cannot assess precisely how robust they are to changes in the testing set. Note that this last remark (robustness) is also applicable to results using conventional features, by definition: the capacity of classifiers to generalize is always demonstrated to a certain extent, determined by the "testing" database. For more information about the systematic comparison of EDS features with conventional features see Pachet and Roy (2007).

From a methodological perspective, the use of advanced machine learning algorithms to classify and analyze animal sounds opens new perspectives for the understanding of animal communication. This study offers only a first illustration of this potential. It is important to stress that the method used in this study is fully automatic (except the segmentation of barks). No information linked with the particular problem of bark classification was included at any stage of the process. This means that the process can be applied indiscriminately to any other audio classification problem. This also guarantees that the process is unbiased that limits the number of potential preconceptions that researchers may introduce into the construction of good descriptors of the data and permits discovery of structure that they may have otherwise missed. This study illustrates how such type of automatically discovered structure may be interpreted in a specific context and how this type of experiment can complement, in a useful manner, other approaches to the study of animal communication. The promising results obtained strongly suggest that the advanced machine learning approaches deserve to be considered as a new relevant tool for ethology.

Acknowledgments The authors are thankful to the members of the MEOE Magyar Mudi Klub for their assistance with the sound recordings. We also wish to thank Celeste Pongrácz for checking our English in this paper. This study was funded by grants from the Hungarian NSF (OTKA) No. T047235.

References

Au WW (1994) Comparison of sonar discrimination: dolphin and an artificial neural network. J Acoust Soc Am 95:2728–2735

Au WW, Andersen LN, Rasmussen R, Roitblat HL, Nachtigall PE (1995) Neural network modeling of a dolphin's sonar discrimination capabilities. J Acoust Soc Am 98:43–50

Bergeron BP (2003) Bioinformatics computing. Prentice-Hall, New York

Bogacz R, Brown MW (2003) Comparison of computational models of familiarity discrimination in the perirhinal cortex. Hippocampus 13:494–524

Burghardt T, Calic J (2006) Analysing animal behaviour in wildlife videos using face detection and tracking. IEEE, NY, Proc Vis Image Signal Process 153:305–312

Burghardt T, Thomas B, Barham PJ, Calic J (2004) Automated visual recognition of individual African penguins. In: Proceedings of Fifth International Penguin Conference, Ushuaia, September 2004

Cabral G, Pachet F, Briot J-P (2005) Automatic X traditional descriptor extraction: The case of chord recognition. In: Proceedings of the 6th international conference on music information retrieval (ISMIR'2005), London, September 2005

Calic J, Campbell N, Calway A, Mirmehdi M, Burghardt T, Hannuna S, Kong C (2005) Towards intelligent content based retrieval of wildlife videos. In: Proceedings of the Sixth WIAMIS, Montreux, April 2005

Cohen JA, Fox MW (1976) Vocalizations in wild canids and possible effects of domestication. Behav Processes 1:77–92

Defréville B, Roy P, Rosin C, Pachet F (2006) Automatic recognition of urban sound sources. In: Proceedings of the 120th AES Conference, Athens

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

Fielding AH (1999) Machine learning methods for ecological applications. Kluwer Academic Publishers, New York

Fitch WT, Neubauer J, Herzel H (2002) Calls out of chaos. The adaptive significance of nonlinear phenomena in mammalian vocal production. Anim Behav 63:407–418

Frommolt K-H, Goltsman ME, MacDonald DW (2003) Barking foxes, *Alopex lagopus*: field experiments in individual recognition in a territorial mammal. Anim Behav 65:509–518

Gerhardt HC (1992) Multiple messages in acoustic signals. Semin Neurosci 4:391–400

Hatzivassiloglou V, Duboué PA, Rzhetsky A (2001) Disambiguating proteins, genes, and RNA in text: a machine learning approach. Bioinformatics 17(Suppl. 1):97–106

Hauser MD (1996) The evolution of communication. MIT Press, Cambridge

Holekamp KE, Boydston EE, Szykman M, Graham I, Nutt KJ, Birch S, Piskiel A, Singh M (1999) Vocal recognition in the spotted hyena and its possible implications regarding the evolution of intelligence. Anim Behav 58:383–395

King RD, Hirst JD, Sternberg MJE (1993) New approaches to QSAR: neural networks and machine learning. Perspect Drug Discov Des 1:279–290

King RD, Muggleton S, Lewis RA, Sternberg MJE (1992) Drug design by machine learning: the use of inductive logic programming to model the structure-activity relationships of trimethoprim analogues binding to dihydrofolate reductase. PNAS 89:11322–11326

King RD, Sternberg MJE (1990) Machine learning approach for the prediction of protein secondary structure. J Mol Biol 216:441–457

Kohavi R, Sommerfield D (1995) Feature subset selection using the wrapper method: Overfitting and dynamic search space topology, First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal

Koza J (1992) Genetic Programming: on the programming of computers by means of natural selection. MIT Press, Cambridge

Kremliovsky M, Kadtke J, Inchiosa M, Moore P (1998) Characterization of dolphin acoustic echo-location data using a dynamical classification method. Int J Bifurcat Chaos 8:813–823

Manser MB, Seyfarth RM, Cheney DL (2002) Suricate alarm calls signal predator class and urgency. Trends Cogn Sci 6:55–57

Mitchell TM (2005) Using machine learning and cognitive modeling to understand the fMRI measured brain activation underlying the representations of words and sentences. In: Proceedings of collaborative research computational neuroscience workshop, Arlington

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

Monceaux J, Pachet F, Amadu F, Roy P, Aymeric Z (2005) Descriptor-based spatialization. In: Proceedings of AES conference 2005, Athens

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

Muggleton S, King RD, Sternberg MJE (1992) Protein secondary structure prediction using logic. Protein Eng 5:647–657

Owren MJ, Rendall D (2003) Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: perceptual experiments with human (*Homo sapiens*) listeners. J Comp Psychol 117:380–390

Obach M, Wagner R, Werner H, Schmidt H-H (2001) Modeling population dynamics of aquatic insects with artificial neural networks. Ecol Modell 146:207–217

Pachet F, Zils A (2004) Automatic extraction of music descriptors from acoustic signals. In: Proceedings of ISMIR 2004, Barcelona

Pachet F, Roy P (2007) Exploring billions of audio features. In: Proceedings of CBMI (Content-Based Multimedia Indexing), Bordeaux, France

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

Pongrácz P, Molnár Cs, Miklósi Á, Csányi V (2006) Human understanding of dog barks: why we did not know what we knew all along? Appl Anim Behav Sci 100:228–240

Recknagel F (2001) Applications of machine learning to Ecol Model. Ecol Modell 146:303–310

Schassburger RM (1993) Vocal communication in the timber wolf, *Canis lupus*, Linnaeus. Advances in ethology (No. 30). Paul Parey Publishers, Berlin

Schleiter IM, Obach M, Borchardt D, Werner H (2001) Bioindication of chemical and hydromorphological habitat characteristics with benthic macro-invertebrates based on artificial neural networks. Aquatic Ecol 35:147–158

Seyfarth RM, Cheney DL, Marler P (1980) Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. Science 210:801–803

Slobodchikoff CN, Kiriazis J, Fischer C, Creef E (1991) Semantic information distinguishing individual predators in the alarm calls of Gunnison's prairie dogs. Anim Behav 42:713–719

Stockwell DRB (2006) Improving ecological niche models by data mining large environmental datasets for surrogate models. Ecol Modell 192:188–196

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

Witten I, Eibe F (1999) Data Mining, Morgan Kaufmann Publishers, San Francisco

Yin S (2002) A new perspective on barking in dogs (Canis familiaris). J Comp Psychol 119:189–193

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

Zhang L, Samaras D, Klein N, Tomasi D, Cottone L, Leskovjan A, Volkow N, Goldstein R (2005a) Exploiting temporal information in functional magnetic resonance imaging brain data, In: Proceedings of MICCAI, Palm Springs, pp 679–687

Zhang L, Samaras D, Tomasi D, Volkow N, Goldstein R (2005b) Machine Learning for clinical diagnosis from functional magnetic resonance imaging. In: IEEE proceedings of CVPR, San Diego, pp 1211–1217