Automatic Classification of Context in Induced Barking

Humberto Pérez-Espinosa[1], José Martín Pérez-Martínez[2], José Ángel Durán-Reynoso[2], and Verónica Reyes-Meza[3]

[1] CONACYT Research Fellow – CICESE-UT3, Tepic, Nay., Mexico

[2] CICESE-UT3, Haramara TIC-LAB, Tepic, Nay., Mexico

[3] Universidad Popular Autónoma del Estado de Puebla, Puebla, Mexico

hperez@cicese.mx, joseperez@cicese.mx, veronica.reyes@upaep.mx
http://idi.cicese.mx/ut3/

Abstract. In this study, we present the results of classification experiments of induced dog barks in different contexts of behaviour. We applied four validation schemes to trained models in order to determine the level of individuals dependency for context classification. We did an analysis based on feature selection techniques to determine the best acoustic low-level descriptors for this task. Results showed that classification performance decreases when the model is evaluated leaving out acoustic information of individuals in the training stage. The acoustic feature set used in our experiments shown better results in comparison with other works using the same data.

Keywords: Barking classification, acoustic characterization, machine learning

1 Introduction

The bark is the most distinctive vocalization of dogs. It occurs very frequently in a wide range of contexts and situations. For humans, many times this chaotically noisy vocalization is annoying. However, people are able to recognize their dogs by their barks [9], to categorize dog barks correctly [12] and even to perceive emotional information from acoustic parameters of dog barks [13]. Some authors say that it has an important function for expression, becoming more and more sophisticated during dog domestication.

Dog barking and other vocalizations of dogs have been studied from different points of view. On one side, researchers have been trying to answer to research questions regarding the function and type of the information carried by dog

barks. For example, in [14] the possible communicative function of dog barking is discussed. Lord et al. [7] explored the functional hypothesis that barking is associated with mobbing behaviour and the motivational states that accompany mobbing.

Other studies have been focused on the acoustic properties of dog vocalizations and mainly on barks. Interesting acoustic patterns have been found from the analysis of the relation between regular and irregular components of the signal. For example, [9] studied the harmonic-to-noise ratio to rank dog vocal utterances from noisy to clear by quantifying the amount irregular energy. Molnar et al. [10] found that individuals are more successfully identified by humans when they listen to a low harmonic-to-noise ratio barks. Some authors have parametrized vocalization of dogs using objective techniques to describe the relationship between sound structure, signal function and social context. For example, in [4] they used sonography to determine the complexity of the dog's vocal repertoire and its communicative value.

Classification of barks based on context has been aboard by some authors. Yin et al. [18] analysed spectrograms of 4,672 barks from 10 dogs generated in 3 different contexts: disturbance, isolation and play. They found specific particularities in frequency and amplitude measurements for each context. Molnar et al. [9], analysed tonality, frequency and intervals between barks produced in 7 different contexts. They tested the ability of human listeners to discriminate between dogs when the context in which bark was recorded changes. For example, they found that for listeners it is easier to recognize the individual dog when barked at a stranger than if they listen when the dog was separated from its owner.

More recently, artificial intelligence techniques have been utilized to automatically classify barks and other dog's vocalizations. In [8] they used a Bayesian classifier for two classification problems, recognition of dogs and categorization of barks into context. They constructed a set of acoustic descriptors using an evolutionary algorithm and feature selection techniques. Larrañaga et al. [6] compared several supervised machine learning methods for four classification tasks: sex, age, context and individual. They tested four machine learning methods and a set of 29 acoustic measures extracted from each barking recording. In the case of context classification, they tested for two learning settings, a single model for all dogs and one model per each dog. Both works were done using a database of Mudi dog barks.

In this work we used the same database of Mudi dog barks previously analysed in the works by Molnar et al. and Larrañaga et al. Our contributions and goals with this work are motivated by two research questions:

1. Which are the best low-level descriptors for barks context classification?
2. How individual dependant is bark context classification?

We analysed the pertinence of a set of low-level acoustic descriptors that has been used for emotion recognition in voice. We implemented a *leave one dog out* validation scheme and compared with other validation schemes to evaluate the accuracy of our models and individual dependency. The main goal with this

project is to train classification models for bark context classification to classify new dogs.

2 Data

We used in our experiments the data collected by Pongrácz et al. [15]. They captured Mudi dog barks, a medium sized Hungarian breed of shepherd dogs. Barks were induced in dogs by performing a predefined protocol of seven different behavioural contexts described below:

1. Alone: The owner and the experimenter take the dog to an outdoor area. The owner leaves the dog tied and walked out of the dog's sight.
2. Ball: The owner holds a ball or toy 1.5m in front of the dog.
3. Fight: The trainer attacks the owner and the dog. The owner keeps the dog on a leash.
4. Food: The owner holds the dogs food bowl 1.5m in front of the dog.
5. Play: The owner plays a game with the dog.
6. Stranger: The experimenter appears at the dog garden or in front of the dog.
7. Walk: The owner behaves as if he/she is preparing for a walk with the dog.

The barks were recorded in a different number of bouts for each dog. With an exception of the contexts Alone and Fight, all recordings were done at the dog's residence. Recordings were made with a tape recorder and a microphone. During recordings, the experimenter stood in front of the dog and faced it while holding the microphone within 1 to 4 meters of the dog. Barks were digitalized with a 16-bit quantization and 22.05 kHz sampling rate. Waveforms were rescaled so that its highest amplitude peak was at -6 dB.

2.1 Annotation and Segmentation

Original recordings were manually segmented at single bark sound level. Segment length ranges approximately from 0.1 to 0.8 seconds. Original recordings and segments are separately stored by dog ID and context. In Fig. 1 (generated by Praat [2]), we can see an original recording that is segmented, eliminating the pause periods and keeping the single bark fragments which are the analysis unit in our experiments.

The data set consist of 6,614 single barks distributed in seven contexts as shown in Table 1. In this same table, it is shown the number of samples of each context used for the different validation methods used in this work. Validation methods are explained in section 5. The barks correspond to 12 dogs as shown in Table 3. As we can see in this table, we have an unbalanced number of samples per class. We can also notice that not all dogs are represented in all contexts. This is due to the complexity of the bark induction protocol implementation. It is a fact that not every dog reacts in the same way and with the same proportion to the stimuli.

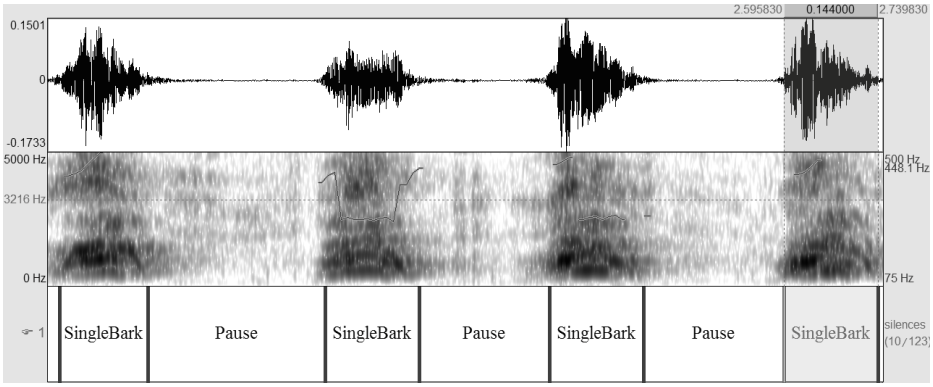

Fig. 1. A recording containing four single bark sounds. Each single bark is stored in a separated audio file.

Table 1. Number of instances used in validation method for each context

OMPD, 10FCV and LODOV	Resample	Context
758	83	Alone
1,004	132	Ball
1,056	143	Fight
833	109	Food
752	119	Play
1,425	226	Stranger
786	100	Walk
6,614	912	Total

3 Acoustic Features Extraction

We used the openSMILE [3] software to extract the following Low-Level Descriptors (LLDs) included in the large openSMILE emotion features set. This acoustic features set was designed for emotion recognition in the human voice.

Melspec N-band Mel / Bark / Semitone - frequency spectrum (critical band spectrum) by applying overlapping triangular filters equidistant on the Mel / Bark / Seminote - frequency scale to an FFT magnitude spectrum.

MFCC The first 12 Mel-frequency Cepstral coefficients are computed on the critical band spectrum.

Energy Computes logarithmic (log) and root-mean-square (RMS) signal energy from PCM frames.

Spectral Bands Computes energy in the given spectral band by summation of FFT bins in the band. The bands computed are 0-250, 0-650, 250-650 and 1000 - 4000.

Spectral Roll Off Compute X*100% spectral roll-off point. The X*100% spectral roll-off point is determined as the frequency below wich X*100% of the total signal energy fall.

Table 2. F-measure for each context obtained in deferent validation settings

OMPD	10FCV	Resample	LODOV	Context
0.93	0.80	0.67	0.30	Alone
0.80	0.67	0.62	0.28	Ball
0.95	0.84	0.81	0.57	Fight
0.82	0.67	0.72	0.31	Food
0.84	0.72	0.59	0.20	Play
0.92	0.79	0.76	0.47	Stranger
0.79	0.64	0.57	0.32	Walk
0.87	0.74	0.69	0.37	Weighted Average

Table 3. Number of instances used in validation for each dog

OMPC, 10FCV and LOCOV	Resample	Dog
275	106	d05
1,007	112	d12
465	115	d14
219	0	d18
968	101	d23
1650	105	d24
693	114	d09
336	131	d16
686	128	d20
124	0	d27
108	0	d10
83	0	d26
6,614	912	Total

Spectral Flux Computes spectral Flux for N FFT bins
Spectral Centroid Computes spectral centroid at time t.
Spectral MaxPos Computes the position of the maximum magnitude spectral bin
Spectral MinPos Computes the position of the minimum magnitude spectral bin.
Voice Prob Computes the probability of voicing via a Cepstrum based method.
F0Env F0 envelope (exponential decay smoothing)
F0 Computes the fundamental frequency via an ACF based method.
ZCR Computes these time signal properties:

LLDs are computed using a frame size of 25 ms and a frame step of 10 ms. A smoothing data contours process is applied by a moving average filter. Delta and double delta regression coefficients are calculated for the values of LLDS in each frame. In order to have the same number of attributes for each single bark recording, regardless of its duration, 39 statistical functions are calculated over the values of the LLDs, its deltas and its double deltas coefficients in each frame

Table 4. F-measure for each dog obtained in deferent validation settings

OMPC	10FCV	Resample	LOCOV	Dog
0.95	0.82	0.89	0.20	d05
0.99	0.97	0.93	0.91	d12
0.97	0.91	0.91	0.65	d14
0.98	0.94	-	0.71	d18
0.98	0.95	0.88	0.86	d23
0.99	0.98	0.94	0.93	d24
0.96	0.94	0.92	0.65	d09
0.95	0.93	0.93	0.73	d16
0.94	0.90	0.87	0.73	d20
0.94	0.87	-	0.67	d27
0.94	0.83	-	0.20	d10
0.94	0.92	-	0.02	d26
0.97	0.94	0.91	0.76	Weighted Average

of the recording. Finally, we obtain a total of 6,552 attributes for each single bark sample. Table 6 shows the number of acoustic features per each LLD.

4 Acoustic Features Selection

After an experimentation stage with several feature selection methods, we decided to use the *Relief Attribute* evaluation method as implemented in Weka [5]. This method as shown the best the best accuracy rates when we took the 500 best-ranked attributes. These features were individually evaluated from the original feature set of 6,552 attributes in order to obtain the best attributes and reduce the dimensionality of the attributes vector. Table 6 shows the number of selected acoustic features per each LLD.

5 Evaluation of Classification of Context and Dog

We used the machine learning technique *Support Vector Machines*(SVM) using a polynomial kernel [5] to classify by context and by dog. We selected SVM given that this technique has shown good results in previous works using a similar acoustic feature set [11]. The validation was made by four methods:

One Model per Dog (OMPD) with the objective of measuring the impact and dependency of individuals in the classification, we implemented a scheme of validation where a classification model is trained with the samples of only one dog. Then the trained model is evaluated by 10FCV. Accuracy statistics is calculated on the accumulated confusion matrix. We included this validation scheme to test the opposite scenario to a dog independent model.

10 Fold Cross Validation (10FCV) In this validation scheme a classification model is trained using the 90% of the samples in the dataset and tested it with the 10% left out. This validation round is repeated 10 times, each time leaving out a different set of samples. We used this validation scheme to have a baseline accuracy. However, given that several samples are extracted from the same recording, they could generate an effect of pseudo-replication.

Resample Dogs with the fewest samples are discarded. We eliminated the four less represented dogs. After this step, we applied the Re-sample method as implemented in [5] to obtain a random subsample of the dataset. We kept the 15% of the samples with a bias to uniform the number of samples from each dog, without replacement of samples. The number of kept samples for each dog is shown in Table 3. Classification accuracy is evaluated using this reduced dataset and 10FCV. We included this validation scheme in order to compare our results with the reported by [6] where they used the same data and similar method for re-sampling.

Leave One Dog Out Validation (LODOV) with the objective of measure the impact and dependency of individuals in the classification. We implemented a scheme of cross validation where a classification model is trained using all the samples of N-1 dogs and tested it with the one left out. Where N is the total number of dogs in the data set, 12 dogs in our case. This validation round is repeated N times, each time leaving out a different dog. Accuracy statistics is calculated on the accumulated confusion matrix.

Table 2 shows the results, in terms of F-measure, of automatic classification per class. F-measure is a classification performance metric that is calculated as the harmonic mean of precision and recall. It may be obvious to expect that any setting that mixes dogs identities during training is going to have a better classification performance but, it is important for the goals of this work to have a clear idea of how big is the impact. We can see that there is a significant difference in classification performance depending on the evaluation scheme. While we can observe an excellent performance for OMPD relatively good performance for 10FCV and Resample, for LODOV we obtained a low performance. Classification performance per context was similar in the four evaluation schemes. Fight was the context with the best results in the four schemes followed by Stranger. The contexts with the lowest performance were Ball, Walk and Play.

When we evaluated by Resample, we obtained an F-measure of 0.69 (as shown in Table 2) and an accuracy of 68.64%. This represents an improvement on the results reported by Larrañaga et al. [6] using the same data and the same evaluation scheme. They obtained an accuracy of 55.50% using a k-nearest neighbour classifier and a wrapper feature selection method. The acoustic features they used were mainly spectral energy and voice cycle measurements.

Table 4 shows the results for individuals classification. Even when dog identification is not the main target of this work, this experiment is important to illustrate that the barks of each dog have evident acoustic particularities regardless the context. It is a fact consistent with contextual plasticity, the extent to which the behaviour of a given animal varies across contexts [17].

We evaluate dog classification with the same four validation schemes used in the previous experiment. For the third scheme, we leave one context out instead of one dog. As we can see, in general, it was an easier classification task. Dogs were classified with a high accuracy except the ones with few samples.

6 Context Grouping

In Table 5 we show the results of automatic classification when grouping contexts. We defined some group according to Arousal and Valence, which are frequently used as human emotion descriptors [16]. Arousal is the level of awakeness or reactivity to stimuli.Valence is the intrinsic attractiveness (positive) or aversiveness (negative) of an event. Contexts were grouped in the following way:

Experiment 1 Negative Valence (Fight, Stranger, Alone) vs Positive Valence (Walk, Ball, Play, Food)

Experiment 2 High Arousal (Fight, Stranger, Walk, Ball, Play) vs Low Arousal (Alone, Food)

Experiment 3 Negative Valence and High Arousal (Fight, Stranger) vs Positive Valence and High Arousal (Walk, Ball, Play) vs Low Arousal (Alone, Food)

Table 5. F-measure for each dog obtained in contexts grouping

Groups	10FCV	LODOV
Experiment 1	0.85	0.71
Experiment 2	0.85	0.72
Experiment 3	0.78	0.58

We performed these experiments to test the acoustic similarities among barks according to the probable emotional state. Table 5 shows that the criteria used to group barks allowed to obtain a relatively good classification performance.

7 Acoustic Features Analysis

Table 6 shows the results of our analysis on acoustic features classification performance. This table shows the number of features originally extracted and also shows the number of selected features for each LLD. The F-measure for each LLD was calculated by group and individually. As mentioned above, we extracted 5,552 features from each single bark. These features are organized into six LLD groups. We tested the performance of LLDs by group and individually to have a better understanding of the discrimination capabilities of these acoustic descriptors. These results were obtained by evaluating separately the features set

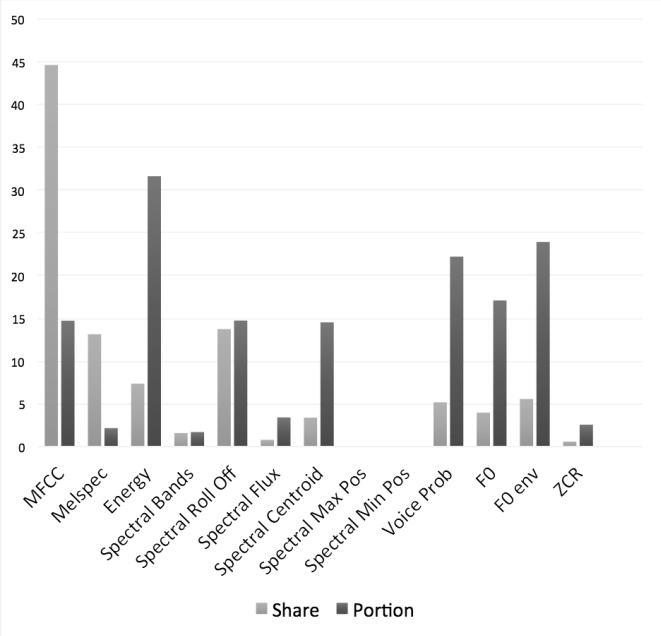

Fig. 2. Share and Portion metrics computed for each LLD.

from each LLD. The validation scheme for this experiment was 10FCV. MFCC was the best LLD. Using only these features we obtained an F-Measure of 0.69, while using all LLDs we obtained an F-Measure of 0.74, as can be seen in Table 2. Melspec was the second best LDD. Energy, Spectral Roll Off and Spectral Centroid also showed an important contribution. On the other hand, there were LLDs that did not provide information, such as Spectral Max Pos and Min Pos.

In Fig. 2 we plot Share and Portion which are measures proposed in [1] to assess the impact of different types of features on the performance of automatic recognition.

Share shows the contribution of each LLD to the selected set of acoustic features. It is computed as the percentage of selected features of one LLD from the total number of features in the selected feature set.

Portion shows the contribution of each LLD weighted by the number of features per type. It is computed as the percentage of selected features of one LLD from the number of features of that LLD included in the original feature set.

As we can see MFCC is the LDD with highest Share. The 45% of the selected features belong to MFCC group. This amount of MFCC features represent the 15% of the MFCC features originally included. Energy is the LLD with the highest Portion The 7% of the selected features belong to Energy group. This amount of Energy features represent the 32% of the Energy features originally included.

Table 6. Number of features for each LLD and number of selected features. F-measure for LLD calculated in groups and individually.

LLD Group	F-Measure	LLD	F-Measure	SelFeatures	F-Measure	OrigFeatures
MFCC	0.69	-	-	224	0.69	1,521
Melspec	0.48	-	-	66	0.67	3,042
Energy	0.41	-	-	37	0.44	117
Spectral	0.47	Bands	0.21	8	0.34	468
		Roll Off	0.41	69	0.40	468
		Flux	0.18	4	0.38	117
		Centroid	0.33	17	0.40	117
		Max Pos	-	0	0.35	117
		Min Pos	-	0	0.26	117
Pitch ACF	0.38	VoiceProb	0.25	26	0.34	117
		F0	0.18	20	0.18	117
		F0 env	0.18	28	0.18	117
		ZCR	0.17	3	0.34	117

Melspec is an interesting case. This LLD have a good performance for classification as shown in Table 6 even when its representation in the selected set is not as significant as other LLDs (Share 13.1%, Portion 2.2%).

8 Conclusions

From the results obtained we can conclude that there is a high dependency on individuals in context classification of barks. In other words, each dog shows a particular way to bark in each context. We saw that context recognition when building models for each individual have very good results, 80% F-measure or higher for all contexts. On the other hand, when we leave one dog out of the training, and then use its samples to test the model, F-measure is not higher than 0.57%. The more dog specific was the evaluation the better the classification performance was.

Dog recognition seems to be an easier classification task than context classification. We obtained good classification performance even when the classification models were evaluated leaving one context out. This mean that a dog can be recognized among other dogs by its barking regardless the context of barking induction.

We were able to corroborate that MFCC, a widely used LLD for human voice analysis mainly speech and speaker recognition, is a good acoustic descriptor to bark context classification task. Using only this descriptor, it is possible to characterize dog barks and build classifiers with a similar performance than classifiers built with a much larger set of descriptors. Energy is also a good LLD for bark classification. This type of acoustic feature provided a high portion of features to the selected set. Melspec features are able to characterize dog barks using a relatively low share and portion from the original number of features.

An interesting result was obtained when dog bark contexts were grouped by Valence and Activation, two primitives used for human emotions modelling. We saw that barking could be analysed in terms of emotion-related information.

Acknowledgments. This research work has been carried out in the context of the "Cátedras CONACyT" programme funded by the Mexican National Research Council (CONACyT).

References

1. Batliner, A., Steidl, S., Schuller, B., Seppi, D., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Aharonson, V., Kessous, L., et al.: Whodunnit–searching for the most important feature types signalling emotion-related user states in speech. Computer Speech & Language 25(1), 4–28 (2011)
2. Boersma, P., Weenink, D.: Praat: doing phonetics by computer [computer program]. (2013)
3. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: the munich versatile and fast open-source audio feature extractor. In: Proceedings of the international conference on Multimedia. pp. 1459–1462. ACM (2010)
4. Feddersen-Petersen, D.: Vocalization of european wolves (canis lupus lupus l.) and various dog breeds (canis lupus f. fam.). Archiv fur Tierzucht 43(4), 387–398 (2000)
5. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD explorations newsletter 11(1), 10–18 (2009)
6. Larranaga, A., Bielza, C., Pongrácz, P., Faragó, T., Bálint, A., Larranaga, P.: Comparing supervised learning methods for classifying sex, age, context and individual mudi dogs from barking. Animal cognition 18(2), 405–421 (2014)
7. Lord, K., Feinstein, M., Coppinger, R.: Barking and mobbing. Behavioural processes 81(3), 358–368 (2009)
8. Molnár, C., Kaplan, F., Roy, P., Pachet, F., Pongrácz, P., Dóka, A., Miklósi, Á.: Classification of dog barks: a machine learning approach. Animal Cognition 11(3), 389–400 (2008)
9. Molnár, C., Pongrácz, P., Dóka, A., Miklósi, A.: Can humans discriminate between dogs on the base of the acoustic parameters of barks? Behavioural processes 73(1), 76–83 (2006)
10. Molnár, C., Pongrácz, P., Faragó, T., Dóka, A., Miklósi, Á.: Dogs discriminate between barks: the effect of context and identity of the caller. Behavioural processes 82(2), 198–201 (2009)
11. Pérez-Espinosa, H., Reyes-García, C.A., Villaseñor-Pineda, L.: Acoustic feature selection and classification of emotions in speech using a 3d continuous emotion model. Biomedical Signal Processing and Control 7(1), 79–87 (2012)
12. Pongrácz, P., Molnár, C., Dóka, A., Miklósi, Á.: Do children understand man's best friend? classification of dog barks by pre-adolescents and adults. Applied animal behaviour science 135(1), 95–102 (2011)
13. Pongrácz, P., Molnár, C., Miklósi, Á.: Acoustic parameters of dog barks carry emotional information for humans. Applied Animal Behaviour Science 100(3), 228–240 (2006)

14. Pongrácz, P., Molnár, C., Miklósi, Á.: Barking in family dogs: an ethological approach. The Veterinary Journal 183(2), 141–147 (2010)
15. Pongrácz, P., Molnár, C., Miklósi, A., Csányi, V.: Human listeners are able to classify dog (canis familiaris) barks recorded in different situations. Journal of Comparative Psychology 119(2), 136 (2005)
16. Scherer, K.R.: Psychological models of emotion. The neuropsychology of emotion 137(3), 137–162 (2000)
17. Stamps, J., Groothuis, T.G.: The development of animal personality: relevance, concepts and perspectives. Biological Reviews 85(2), 301–325 (2010)
18. Yin, S., McCowan, B.: Barking in domestic dogs: context specificity and individual identification. Animal Behaviour 68(2), 343–355 (2004)