Detection of Basic Emotions from Cats' Meowing

Qianlong Shou, Yumeng Xu, Junjun Jiang, Min Huang(✉), and Zhongzhe Xiao(✉)

School of Optoelectronic Science and Engineering, Soochow University, Suzhou 215006, Jiangsu, China
{hmin,xiaozhongzhe}@suda.edu.cn

Abstract. Basic emotional states in valence sense as positive, neutral, and negative are studied with automatic classification on cats' meowing signals, aiming to help human-cat interaction and human emotion regulation by pets keeping. The ground truth of meowing samples is marked by subjective evaluation from multiple raters with the help of cats' facial expression, body movement, and interaction with cat owners in video clips. Acoustic features extracted from voice energy, zero crossing rate, and MFCC are proved to be effective in cats' emotion recognition. The highest accuracy reaches 97.40% on selected best feature subset with LogitBoost model.

Keywords: Cats emotions · Acoustic features · Recognition

1 Introduction

Voice, as an effective communication style, plays an essential role in the expression of feelings. In recent years, researchers have yielded numerous remarkable results in the emotional analysis of speech for humans [1–3] and various speech emotion datasets [4] have been obtained. Furthermore, the recognition for emotions of human has achieved a much higher accuracy. Overall, great progress has been made in the study of emotions in human voice. However, few studies have focused on the analysis of emotions for animals, and there are a few affective computing techniques to recognize the emotions for mammals except humans.

Some researchers have studied the barking of dogs and analyzed the emotions contained in dog barkings. Acoustic characteristics have been discussed for the recognition of dogs by their barkings [5] and many features have been proposed, as well as methods, while emotions of cats, who are also important accompany pets of humans, are not yet studied thoroughly with automatic analysis. A good model in recognition of cats' emotions, will greatly help human, especially new owners of cats, to quickly develop a better interaction with their pets, and make the most advantage of cat keeping, for accompany, or even emotion regulation for human (cat owner).

X. Shao et al. (Eds.): CSMT 2020, LNEE 761, pp. 149–160, 2021.
https://doi.org/10.1007/978-981-16-1649-5_13

From the related studies on human emotions expressed by voice, one existing problem is that the emotion categories never reached any universal agreement. Relatively commonly accepted emotion taxonomies include Ekman's "big six" [6], or two-dimensional model with valence and arousal [7]. Application dependent definition of emotion categories is also a common manner, such as in the case of several widely used emotional speech datasets [8–10]. Although there are currently very few studies on cats' emotions, similar investigation has been made with dogs as behaviour and emotion models of companion robots [5,11–13]. For example, application dependent emotions as happiness, despair, fear, anger, and surprise are used in [12]. In this work on cats' emotions by meowing voice, we choose to use a simple way as the starter, with three states in valence sense as positive, neutral, and negative, to describe the cats' most basic emotions.

In the machine learning based approaches, a dataset with reliable labeling of ground truth to each sample is the essential basis. For example, in the work with dog barkings, perception tests indicate that acoustic parameters, including tonality, pitch and inter-bark time intervals, are strongly related with emotions and affect greatly on listeners' judgment [14–17]. In building meowing dataset in this work, subjective evaluation with human judgement will also be used, while with the help of video contents including cat facial expressions, body movement, *etc.*., because human judgement of cats' emotions only by meowing voice is not a practical activity for most persons.

The rest of the paper is organized as follows. Section 2 introduces the construction process of the dataset of cats' meowing. Section 3 describes the process of emotional feature extraction and conducts the feature dimensionality reduction to avoid curse of dimensionality. Section 4 gives a set of experiments for evaluation. Finally, Sect. 5 concludes the paper and presents the future work.

2 Dataset Construction of Cats' Meowing

We aim to perform an automatic detection of cat's basic emotions from cats' meowing in a data driven manner. Thus, to collect a dataset with sufficient cats' meowing samples with reliable labelling is an essential preparation. The collected meowing samples can then be regarded as cats' "language" in the cat emotion detection.

There are two basic concerns in the construction of this dataset. First, meowing samples from only one or two cats will introduce great influence of the cat individual, and the common clues in cats' emotions expressing by voice cannot be fully discovered. Second, with only the audio signal of cats' meowing, we cannot accurately judge the cats' emotional states. The cats' facial expressions, body movements, and the surrounding situations including their interactions with their owners will help greatly for the judgement. For the above two reasons, video clips from public websites/apps are chosen as the resource of cats' meowing samples in this work. The resource websites we used in data collection include Iqiyi, Tencent Video, Bilibili, Haokan Video, Wesee Video, etc. A lot of "cat persons" are sharing their daily interactions with their cats, with labels or even detailed explanations

to the shared scenes. These sharings facilitate persons who like "cyber cat petting" to satisfy their cataholic. By collecting samples from these video websites, we can obtain cats' scenes from a large number of different cats. The labels and explanations also make our data collection much easier, because the owners of the cats know their cats very well and the labels can be seen to be reliable reflecting the cats' emotional states. The construction process of the dataset is illustrated in Fig. 1, from the video clips with owners' labels and explanations.

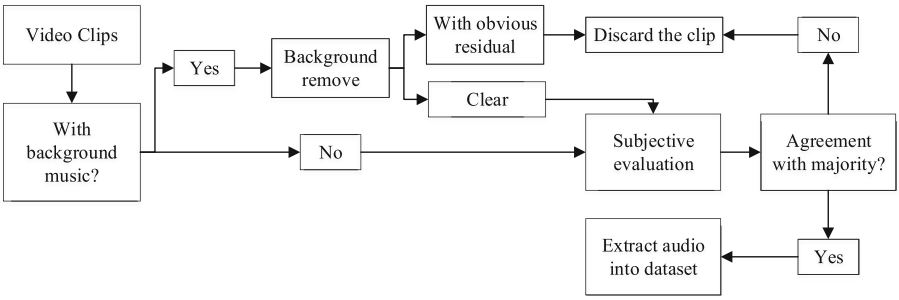

Fig. 1. Flow chart of constructing cats' meowing dataset

One of the problems in the meowing samples collection is that cat owners usually add background music with their uploaded videos, while the background music will significantly influence the analysis of cats' voice. In this case, we first make a preprocessing with Adobe Audition to remove it. If the obvious residual of background can still be heard after the removal, the video clip will be discarded. Video clips with clear removal of background music, together with clips without any background music, are sent to subjective evaluation in the next step.

In subjective evaluation, several raters were asked to evaluate the emotional states of cats. To make the evaluation not too dispersed, we only set three categories of emotions in valence sense as positive, neutral, and negative. Positive emotions include happy, contentment when they get food or play with their owners, sometimes the cat will make snoring like sound to express their satisfaction. Negative emotions include the states such as hunger, scared, anger, etc. The usual states are regarded as neutral. The raters make the evaluations by watching videos, including the cats' facial expressions, body movements, and interactions with cat owners as their basis of judgement, and if the cat owners provided labels or explanation, this will also be very important evidence for the raters. Examples of cats' facial expressions are shown in Fig. 2. Not all raters have to evaluate all collected video clips, but we ensured that each clip received evaluations from at least 3 raters. When majority of raters give consistent judgement, this clip will be marked with the corresponding positive, neutral, or negative label as the ground truth, and the audio part is extracted into the dataset. If no majority judgement exists, the clip will be discarded.

Totally 566 samples are kept in our collected dataset, with 179 positive samples, 141 neutral samples, and 246 negative samples.

(a) Positive (b) Negative (c) Neutral

Fig. 2. Examples of cat emotions presented by facial expressions

3 Feature Extraction

Effective features that can express the characteristics of cat emotions from meowing voice are one of essential factors in machine learning based recognition. Currently, there are very few studies on feature analysis of cat meowing for emotions, we proposed in this paper to adopt experience from emotion recognition works on speech, music, or other common audio signals. A good choice is to use the feature set provided by the challenges of INTERSPEECH, such as the emotion challenge in 2009 [18], or more comprehensive paralinguistics challenges in 2010 and 2013 [19,20]. These feature sets have been proved in a number of work concerning human speech emotion [21], and could be a good starter for this investigation of cat meowing emotion.

Concerning that we only collected several hundreds samples of emotional cats' meowing, which only form into a small scaled dataset, high dimensional feature sets will cause the problem of overfitting. In order to minimize the impact, the feature set from INTERSPEECH 2009 emotion challenge, which is with the fewest dimension of features in this series of challenge feature sets, is adopted in this work. Three categories of features, as prosody features, sound quality features and spectral features, are contained in this feature sets. The overall extraction of these features to apply 12 statistical functions on 16 low-level descriptors (LLDs) and their first order difference, to result into features, as shown in Fig. 3.

The 16 LLDs are zero-crossing-rate (ZCR), root mean square (RMS) energy, fundamental frequency (F0), harmonic-noise ratio (HNR), and first 12 Mel frequency cepstrum coefficients (MFCCs). The functions to be applied on these LLDs range from first order to higher order statistics, including mean, standard

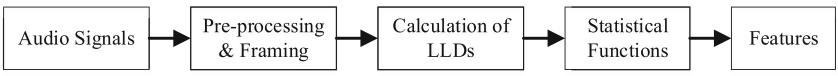

Fig. 3. Feature extraction process

deviation, kurtosis, skewness, maximum and minimum value, relative position, range, and offset and slope of linear regression, together with their mean square error.

The extraction of the above feature set for cats' meowing analysis is based on TUM's open-source openSMILE toolkit [22], with the configuration "emotion_IS09.conf".

4 Automatic Emotion Detection of Meowing

4.1 Experiment Settings

The machine learning approaches for the cats' meowing emotion detection are implemented on WEKA platform [23]. Logistic Regression for classification is chosen as the most basic algorithm in this investigation. As the cats' emotion is not linear, this generalized linear model may not fully present the distinguishing ability of the features, two higher level classifiers based on logistic are used for better performances, as LogitBoost (as in WEKA platform), which uses boosting method based on logistic with maximum likelihood for optimization, and LMT, which builds a tree structure classifier with each node as a logistic model. Beside LMT tree model, another tree model, Random Forest, is also evaluated for comparison.

There is a problem in the collected meowing dataset that the number of samples is extremely unbalanced in each category. This imbalance will significantly influence the reliability of the trained models [24]. Thus, we desampled the negative and positive states, to leave only 141 samples in each category to balance with neutral samples. All the evaluations of models are implemented with 10-fold cross validation, to minimize the bias in dividing such small scaled dataset into training set and test set.

4.2 Classification Results of Cats' Meowing

Automatic classification of cats' emotions from meowing voice is performed with the above mentioned four classifiers, and we present the results in terms of accuracy, AUC (area under ROC curve), and confusion matrices.

The accuracies and kappa statistics from the 4 selected classifiers are compared in Fig. 4. The most basic classifier, logistic, presents relatively poor performance with accuracy of only 63.36%, and AUC as low as 0.56, indicates unreliable emotion detection ability with this method. The compound methods based on logistic, LogitBoost and LMT, get significantly improved accuracies of 94.33% and 86.52%, with AUC of 0.88 and 0.75, respectively. These improvements show that the compound methods fit the cats' emotion detection problem better than the basic logistic methods, and the performance especially benefits from the boosting approach, while the tree structure also helps to get better classification in this task of cats' emotion detection from meowing voice. Another tree based

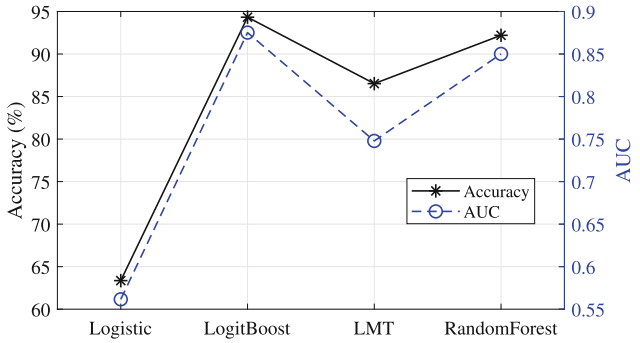

Fig. 4. Accuracies of cats' emotion classification with 4 classifiers

method evaluated here is the Random Forest algorithm, which achieved accuracy of 92.20% with AUC of 0.85. This result is close to that of LogitBoost, and is also highly reliable with high kappa value.

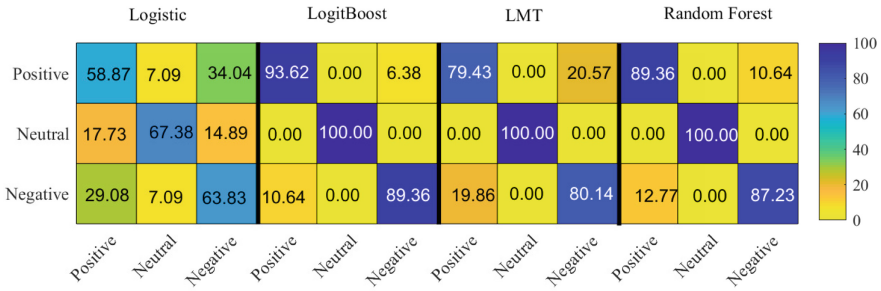

Fig. 5. Confusion matrices of cats' emotion classification (%)

The confusion matrices from the 4 classifiers are shown in Fig. 5, darker colors correspond to higher rates. The worst one, Logistic, presents high confusion to positive or negative from all categories, almost symmetric with positive and negative. A notable phenomenon appears in all other better classifiers that the neutral state is always perfectly classified, and all confusions appear between positive state and negative state. This can be explained by a known fact from human speech emotion that the emotions are easier to be distinguished in arousal dimension than in valence dimension. In this evaluation of cats' emotion from meowing voice, the positive and negative states are defined in valence dimension, while both states present higher arousal than neutral state, thus it leads to the result that the neutral is better recognized than both positive and negative, rather than presented as a middle state between them.

4.3 Further Analysis

From the accuracies of over 90% in cats' emotion classification from LogitBoost and RandomForest, we assume that the cats' emotional states can be detected by the meowing voice, and can be well presented by the INTERSPEECH 2009 emotion challenge feature set. In this subsection, we further analyze the 3 categories of cats' emotion as positive, neutral, and negative with the properties of meowing signals and features.

Frequency Domain Analysis - Spectrum and Spectrogram. Frequency domain properties of cats' meowing voice are displayed in Fig. 6 in form of short time spectrum, from selected typical meowing samples. Similar to human voice, cat voice also presents clear peaks in the spectrum as fundamental frequency and its harmonies. We can see from Fig. 6 that meowing in positive state presents less energy in high frequency band (3000 Hz) than neutral and negative states, and the harmonies are clearer. Meowing in negative state presents a lot of high energy frequencies between the harmony peaks, to make the peak pattern somehow in chaos.

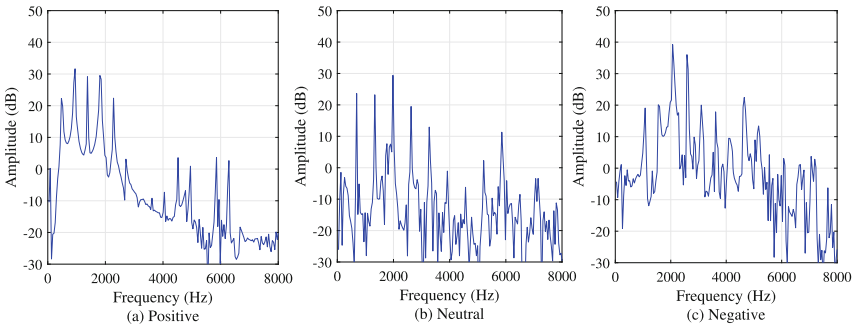

Fig. 6. Spectrum of typical meowing samples of the three categories

A more intuitive illustration of meowing voice can be exhibited in form of spectrogram, as shown in Fig. 7. Cat meowing signals from all 3 emotional states show horizontal stripes in the spectrogram, while the stripes are clearer and thinner in positive state than the other two states, where in neutral states, the stripes are still clear from each other, while in negative state, some of the stripes get blurred together. Neutral state shows smoother F0 trace with the calm voice, positive and negative states that with higher arousal will introduce more fluctuate in F0 trace. Another phenomenon to be noticed is in long scale time domain that, when a cat is in negative state, it tends to produce a longer meowing.

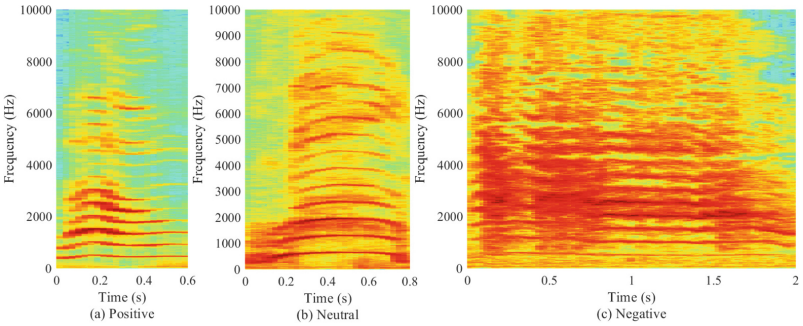

Fig. 7. Spectrogram of typical meowing samples of the three categories

Feature Analysis and Dimension Reduction. This work suffers from a problem that the meowing samples we collected from the internet videos are not sufficiently enough. Even we choose a relatively small scaled feature set, the audio samples in each category are still less than the number of features in the set. Thus, a feature selection, or a feature dimension reduction is necessary for the reliable of this investigation on cats' emotions. In a filter approach of feature selection [25], we ranked the 384 features in INTERSPEECH 2009 feature set in sense of information gain ratio. With from 1 to 30 "good" features, we repeated the automatic classification with the best classifier in Sect. 4.2, LogitBoost. The accuracies are plotted in Fig. 8. The accuracy can reach over 90% with only 4 best features, and increase to over 95% with 8 features. With no less than 10 features, the accuracy stay relative stable between 96% and 97%, where the highest accuracy appears with 17 or 18 features as 97.40%.

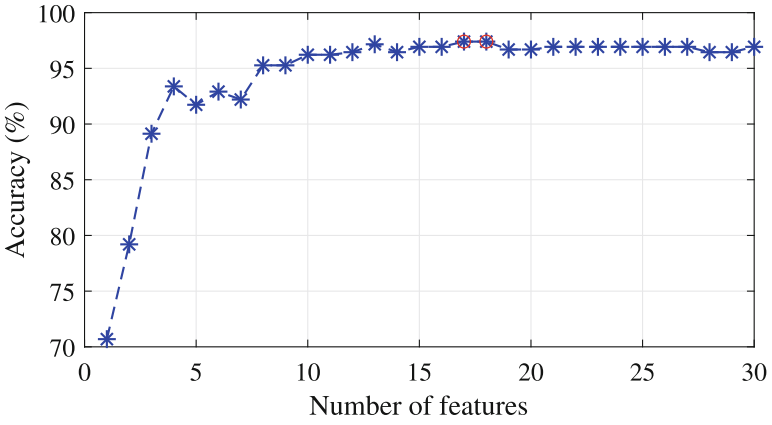

Fig. 8. Accuracy with different number of ranked features

The confusion matrix with the highest accuracy is displayed in Fig. 9. The most significate confusion in this case is that samples of negative state are misjudged as positive at a rate of 4.96%.

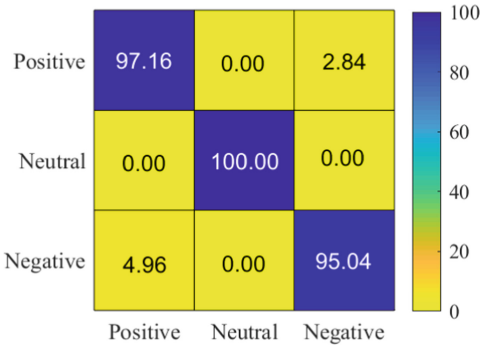

Fig. 9. Confusion matrix with 17 best features

The distribution of 10 selected features is displayed in Fig. 10. As some of the features have similar distribution to each other, these 10 features are not precisely the best 10 features in the ranking. It is shown that these features present different ranges on the three emotional categories, and thus provide distinguishing ability in automatic emotion classification.

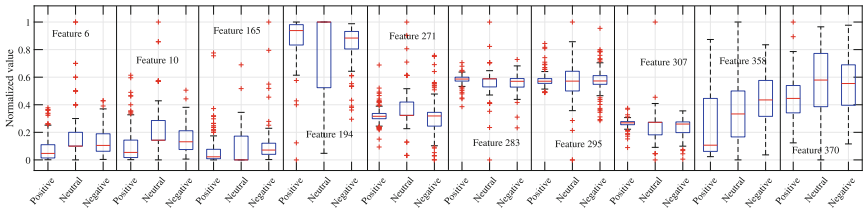

Fig. 10. Distribution of several selected features

The related low level parameters (LLDs as described in INTERSPEECH 2009 feature set) of these features are listed in Table 1. Parameters as RMS energy, ZCR, and MFCC are all important expressive parameters in cats' meowing emotions.

Table 1. Related parameters of the selected features

Feature index	Related LLD	Feature index	Related LLD
6	RMS energy	283	7^{th} MFCC
10	RMS energy	295	8^{th} MFCC
165	ZCR	307	9^{th} MFCC
194	RMS energy	358	ZCR
271	6^{th} MFCC	370	HNR

5 Conclusion

Three categories of cats' emotion, positive, neutral, and negative, as evaluated in valence sense, are investigated with automatic classification on voice signals of cats' meowing. Only audio signals are considered in the learning models, but the ground truth of each sample is determined from video clips with cat voice, facial expression, body movement, as well as their interaction with their owners. Feature set adopted from INTERSPEECH 2009 emotion challenge is proved to be also effective in cats' emotion recognition, and the most expressive features relate to RMS energy, ZCR, and MFCC. The best classification accuracy is obtained from LogitBoost model as 97.40%.

Larger meowing dataset and more detailed emotional categories will be studied in the near future to provide a more accurate and more practical recognition of cats' emotions. Both the aims of this work and the future work focus on the helping of a higher quality human-cat interaction, and make the most of the accompanying role of pets in human psychological adjustment.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Project 61906128 and Project 61802272, in part by the National Natural Science Foundation of Jiangsu Province under project BK20180834.

References

1. Li, S., Yan, Z., Wu, X., Li, A., Zhao, B.: A method of emotional analysis of movie based on convolution neural network and bi-directional LSTM RNN. In: 2017 IEEE Second International Conference on Data Science in Cyberspace (DSC), Shenzhen, pp. 156–161 (2017). https://doi.org/10.1109/DSC.2017.15
2. Xiao, Z., Wu, D., Zhang, X., Tao, Z.: Speech emotion recognition cross language families: Mandarin vs. western languages. In: 2016 International Conference on Progress in Informatics and Computing (PIC), Shanghai, pp. 253–257 (2016). https://doi.org/10.1109/PIC.2016.7949505
3. Kaur, R., Joshi, A.: A study of speech emotion recognition methods. Int. J. Comput. Sci. Mob. Comput. **2** (2013)
4. Xiao, Z., Chen, Y., Dou, W., Tao, Z., Chen, L.: MES-P: an emotional tonal speech dataset in mandarin with distal and proximal labels. IEEE Trans. Affective Comput. (2019). https://doi.org/10.1109/TAFFC.2019.2945322

5. Hantke, S., Cummins, N., Schuller, B.: What is my dog trying to tell me? the automatic recognition of the context and perceived emotion of dog barks. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, pp. 5134–5138 (2018). https://doi.org/10.1109/ICASSP.2018.8461757
6. Ekman, P.: An argument for basic emotions. In: Cognition and Emotion, vol. 6 (1992). https://doi.org/10.1080/02699939208411068
7. Scherer, K.: Psychological models of emotion. In: The Neuropsychology of Emotion (2000)
8. Engberg, I.S., Hansen, A.V., Andersen, O.K., Dalsgaard, P.: Design, recording and verification of a Danish emotional speech database. In: European Conference on Speech Communication and Technology, Rhodes, Greece (1997)
9. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F., Weiss, B.: A database of German emotional speech. In: INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal (2005)
10. Busso, C., Parthasarathy, S., Burmania, A., Abdelwahab, M., Sadoughi, N., Provost, E.M.: MSP-IMPROV: an acted corpus of dyadic interactions to study emotion perception. IEEE Trans. Affect. Comput. **8**, 67–80 (2017). https://doi.org/10.1109/TAFFC.2016.2515617
11. Lakatos, G.: Dogs as behavior models for companion robots: how can Human-Dog interactions assist social robotics? IEEE Trans. Cogn. Dev. Syst. **9**, 234–240 (2017). https://doi.org/10.1109/TCDS.2016.2552244
12. Lakatos, G.: Dogs as behavior models for companion robots: How can Human-Dog interactions assist social robotics? IEEE Trans. Cogn. Dev. Syst. **9**, 234–240 (2017). https://doi.org/10.1109/TCDS.2016.2552244
13. Molnár, C.: Classification of Dog barks: a machine learning approach. Animal Cogn. **11**, 389–400 (2008). https://doi.org/10.1007/s10071-007-0129-9
14. Pongracz, P., Molnar, C., Miklosi, A., Csányi, V.: Human listeners are able to classify dog (Canis familiaris) barks recorded in different situations. J. Comp. Psychol. 119, 136. Washington, D.C (2005). https://doi.org/10.1037/0735-7036.119.2.136
15. Molnár, C., Pongrácz, P., Dóka, A., Miklósi, A.: Can humans discriminate between dogs on the base of the acoustic parameters of barks? Behav. Process/ **73**, 76–83 (2006). https://doi.org/10.1016/j.beproc.2006.03.014
16. Pongrácz, P., Miklósi, D., Csányi, V.: Owner's beliefs on the ability of their pet dogs to understand human verbal communication: a case of social understanding. Curr. Psychol. Cogn. (2000)
17. Faragó, T., Takács, N., Miklósi, A., Pongracz, P.: Dog growls express various contextual and affective content for human listeners. R. Soc. Open Sci. **4** (2017). https://doi.org/10.1098/rsos.170134. England
18. Schuller, B., Steidl, S., Batliner, A.: The interspeech 2009 emotion challenge. In: Proceedings of Interspeech, pp. 312–315 (2009)
19. Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, F., Muller, C., Narayanan, S.: The interspeech 2010 paralinguistic challenge. In: Proceedings of the 11th Annual Conference of the International Speech Communication Association, pp. 2794–2797 (2010)
20. Schuller, B., et al.: The interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, Lyon, France, pp. 148–152 (2013)

21. Deb, S., Dandapat, S., Krajewski, J.: Analysis and classification of cold speech using variational mode decomposition. IEEE Trans. Affect. Comput. **11**, 296–307 (2020). https://doi.org/10.1109/TAFFC.2017.2761750
22. Eyben, F., Wollmer, M., Schuller, B.: Opensmile - the munich versatile and fast open-source audio feature extractor. In: ACM MM, pp. 1459–1462 (2010). https://doi.org/10.1145/1873951.1874246
23. Hall, M.A., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explor. **11**, 10–18 (2008). https://doi.org/10.1145/1656274.1656278
24. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**, 1263–1284 (2009). https://doi.org/10.1109/TKDE.2008.239
25. Kojadinovic, I., Wottka, T.: Comparison between a filter and a wrapper approach to variable subsetselection in regression problems (2000)