









CatMeows: A Publicly-Available Dataset of Cat Vocalizations

Luca A. Ludovico¹, Stavros Ntalampiras¹, Giorgio Presti¹,
Simona Cannas², Monica Battini³, and Silvana Mattiello³

¹ Department of Computer Science, University of Milan, Milan, Italy
{luca.ludovico,stavros.ntalampiras,giorgio.presti}@unimi.it

² Department of Veterinary Medicine, University of Milan, Milan, Italy
simona.cannas@unimi.it

³ Department of Agricultural and Environmental Science, University of Milan,
Milan, Italy
{monica.battini,silvana.mattiello}@unimi.it

Abstract. This work presents a dataset of cat vocalizations focusing on the meows emitted in three different contexts: brushing, isolation in an unfamiliar environment, and waiting for food. The dataset contains vocalizations produced by 21 cats belonging to two breeds, namely Maine Coon and European Shorthair. Sounds have been recorded using low-cost devices easily available on the marketplace, and the data acquired are representative of real-world cases both in terms of audio quality and acoustic conditions. The dataset is open-access, released under Creative Commons Attribution 4.0 International licence, and it can be retrieved from the Zenodo web repository.

Keywords: Audio dataset · Audio signal processing · Bioacoustics · Cat vocalizations · Meows

1 Introduction

Domestic cats (*Felis silvestris catus*) produce different vocalizations, with closed mouth (e.g., purrs and trills), with the mouth held open (e.g., spitting or hissing), and while the mouth is gradually closing or opening (i.e. meows) [2]. Meows are considered as the most relevant type of vocalizations used by domestic cats to communicate with humans [2], whereas undomesticated felids rarely meow to humans [5], despite being common as an intra-specific vocalization [3]. Consequently, the analysis of meows can be useful to investigate cat-human communication.

Concerning widespread companion animals, who are often perceived as social partners [11, 13], few studies have been carried out to understand the mechanisms of their vocal communication with humans. For example, Pongrácz analyzes the effects of domestication on the vocal communication of dogs [23], McComb *et al.* focus on cat purrs [15], and Owens *et al.* propose a visual classification of feral cat vocalizations [18].

In 2019, in the framework of a multi-disciplinary project including veterinarians, psychologists, and experts in sound computing, a study was conducted on the automatic analysis of meows emitted in three different contexts: waiting for food, isolation in an unfamiliar environment, and brushing. The idea was to extract the audio characteristics of the sounds produced in response to heterogeneous stimuli, where the request for food was expected to induce excitement, isolation to cause discomfort, and brushing to provoke either positive or negative responses, depending on the specimen. The results of the project have been published in [17]. After the design phase, the first step of the project was the production of a suitable dataset, called *CatMeows*, now available to the scientific community for research purposes.

The goal of this paper is to describe *CatMeows* from different perspectives, focusing on the protocol used to gather the data (Sect. 2), the internal composition of the dataset (Sect. 3), and some possible application scenarios (Sect. 4). Our conclusions will be summarized in Sect. 5.

As detailed below, *CatMeows* is open access, and it can be retrieved from Zenodo,¹ a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN. The dataset has been assigned the following DOI: [10.5281/zenodo.4007940](https://doi.org/10.5281/zenodo.4007940).

2 Building the Dataset

The goal of this section is to provide details about the whole protocol adopted to obtain the *CatMeows* dataset. Section 2.1 will discuss some design choices, Sect. 2.2 will focus on technical issues related to meow capturing, and Sect. 2.3 will describe post-processing operations.

2.1 Design Choices

For our experiments, one of the critical problems to solve was recording vocalizations while minimizing the potential influences caused by environmental sounds and other noises. Moreover, the characteristics of the room (e.g., its topological properties, reverberation effects, the presence of furniture, etc.) should not affect the captured audio signals. A possible solution would have been to record sounds in a controlled anechoic environment, but, unfortunately, an unfamiliar space was expected to influence the behavior of the cat in the waiting for food and brushing scenarios. In fact, the meowing when placed in an unknown environment, without other stimuli, was one of the reactions our experiment tried to capture and analyse.

Technology in use consists in reliable, commercially available, mass-production sensors, so as to be representative of the sound quality which can be realistically implemented at the time of writing. The use of such an equipment should also ease technological transfer.

¹ <https://zenodo.org/>.

Another constraint regarded the relative position of the microphone with respect to the sound source, that had to be fixed in terms of distance and angle, as many characteristics to be measured in vocalizations could depend on such parameters. As a matter of fact, this requirement prevented the adoption of easy solutions, such as a single omnidirectional microphone or an array of fixed microphones suitably mounted in the room. The input device had to move together with the sound source, keeping the distance and angle as constant as possible.



Fig. 1. Three of the cats participating in the experiment. A Bluetooth microphone is mounted on the collar so as to point upwards and keep an almost fixed distance from their vocal apparatus.

The idea was to capture sounds very close to the source. A number of solutions were explored, including wearable devices somehow attached to the cat's back. Being a whole computing system embedding the recording, processing, and storing chain within a single miniaturized board, such a solution would have been quite cheap, technologically sound, and extremely easy to adopt. Unfortunately, this approach could not be applied to cats, since, according to veterinarians, the perception of a device in close contact with the fur could have produced biased behavioral results. Moreover, the problem of capturing sound in proximity of the vocal apparatus would have persisted.

Thus, a refinement of the original idea was to adopt a very small and lightweight microphone to be placed under the cat's throat through a collar, as shown in Fig. 1. The collar is an object the animal is generally already familiar with, so it was not expected to alter its behavior; in other case, the cat was trained to wear the collar for some days before data recording, until no sign of discomfort (e.g., scratching the collar, shaking, lip licking, yawning, etc.) was reported [7, 8, 16, 19, 20].

Since the placement of the microphone was the same for all cats, sound modifications due to angle, distance and relative position of the mouth were consistent across all recordings, thus marginalizing the effects of these parameters. Some types of information, such as the pitch, did not change, whereas other audio features resulted biased by a constant factor across recordings.

Finally, it could be argued that different situations could alter cats' postures, thus influencing sound production. We did not consider this as a bias,

since we were interested in the final vocalization outcome, regardless of how the cat produced it. A change in posture was considered a legitimate and realistic consequence of the situation. It should be mentioned that if the recording device was able to catch such an aspect, it would have been considered as a useful additional information.

2.2 Capturing Audio Signals

Concerning the input device for audio signals, we turned to Bluetooth connectivity-based products, due to their small dimensions and weight, low power consumption, good transmission range, low price, and acceptable recording quality.

Bluetooth microphones are usually low-budget devices packed with mono earphones and optimized for human voice detection. For example, the frequency range correctly acquired is typically very limited when compared to high-quality microphones. Nevertheless, in this category of devices, there are products presenting all the features required by our experimentation, specifically for the development of automatic classification systems.



Fig. 2. Detailed view of the *QCY Q26 Pro* Bluetooth headset. The microphone is visible in the form of a small hole in the right part of the image.

For the goals of the project, we selected the *QCY Q26 Pro* Bluetooth headset (see Fig. 2), presenting a dynamic range of 98 ± 3 dB. It transmits audio signals via Bluetooth with HFP [25] or HSP [6] profiles, introducing a logarithmic A-law quantization² at a sampling frequency of 8 kHz. As a consequence, the actual range of frequencies we could rely on was 0 to 4 kHz. The fundamental frequency emitted by cats when meowing fall within that range, but some higher-frequency harmonics could be cut.

The size ($15 \times 10 \times 23$ mm) and weight (50 g) of the device were sufficiently small to be carried by a cat without significant behavioral implications. Interestingly, this microphone represented a very low-budget solution: the whole headset, including an earphone (obviously not employed in the experimentation), at the moment of writing can be found in the marketplace for 12 to 25 €.

² The A-law algorithm is a standard used in European 8-bit PCM digital communications systems to optimize the dynamic range of an analog signal for digitizing.

The adoption of the Bluetooth communication protocol greatly simplified the recording chain. In the experimentation, the headset was paired to a smartphone equipped with *Easy Voice Recorder PRO*, a high-quality audio recording application supporting the LPCM³ encoding. This aspect was fundamental to get audio files in a non-supervised environment, e.g. at home, also from people not expert in the IT field, like most cat owners.

2.3 Post-processing

Non-meowing sounds produced by the cats were discarded (13 instances, still available in a separate folder of the dataset), and the remaining sounds were cut so that each file contained a single vocalization, with 0.5 s of leading and trailing silence segments, or *handles*.

Table 1. File naming policy: explanation of the pattern C_NNNNN_BB_SS_00000_RXX.

Placeholder	Description	Values
C	Emission context	B: brushing F: waiting for food I: isolation
NNNNN	Cat name	Cat's unique ID
BB	Breed	MC: Maine Coon EU: European Shorthair
SS	Sex	FI female, intact FN: female, neutered MI: male, intact MN: male, neutered
00000	Cat owner	Cat owner's unique ID
R	Recording session	1, 2 or 3
XX	Vocalization counter	01...99

Files are named after the pattern C_NNNNN_BB_SS_00000_RXX, where each placeholder corresponds to a part of the information carried by the file (see Table 1). Included files follow the `wav` format type.

3 Composition of the Dataset

The cats recruited for the original project described in [17] were 10 adult *Maine Coon* cats (1 intact male, 3 neutered male, 3 intact females and 3 neutered

³ LPCM stands for Linear Pulse-Code Modulation, a standard method to digitally represent sampled analog signals. In a LPCM stream, the amplitude of the analog signal is sampled regularly at uniform intervals, and each sample is quantized to the nearest value within a range of digital linearly spaced steps.

females) belonging to a single private owner and housed under the same conditions, and 11 adult *European Shorthair* cats (1 intact male, 1 neutered male, no intact females and 9 neutered females) belonging to different owners and housed under heterogeneous conditions.

In presence of at least one veterinarian, cats were repeatedly exposed to three different contexts that were expected to stimulate the emission of meows:

1. *Brushing*—Cats were brushed by their owners in their home environment for a maximum of 5 min;
2. *Isolation in an unfamiliar environment*—Cats were transferred by their owners into an unfamiliar environment (e.g., a room in a different apartment or an office). Distance was minimized and the usual transportation routine was adopted so as to avoid discomfort to animals. The journey lasted less than 30 min and cats were allowed 30 min with their owners to recover from transportation, before being isolated in the unfamiliar environment, where they stayed alone for maximum 5 min;
3. *Waiting for food*—The owner started the routine operations that preceded food delivery in the usual environment the cat was familiar with. Food was given at most 5 min after the beginning of the experiment.

The typical vocalization in response to a single stimulus was composed by a number of separated vocalizations. In this case, sound files have been cut according to the strategy explained in Sect. 2.3.

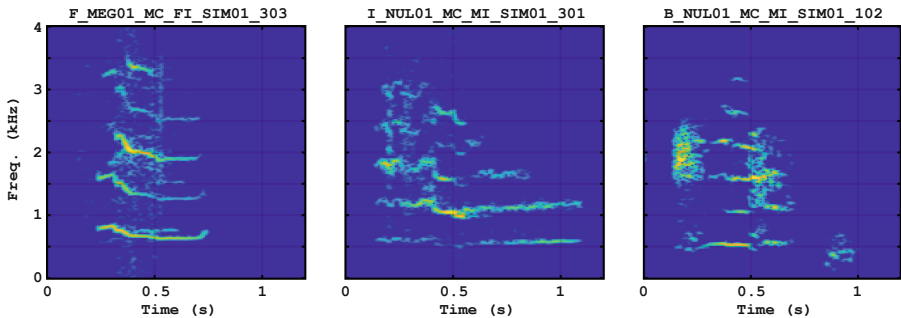


Fig. 3. Time-frequency spectrograms of meows originated by the three considered classes.

The final corpus contains 440 files, organized as shown in Tables 2 and 3. The total number of files referring to *Maine Coon* specimens is 188, and 252 for *European Shorthair* cats. The average length of each file is 1.83 s, with a variance of 0.36 s. About 1 s of each file contains only background noise, due to 0.5 s handles before and after the vocalization.

From the analysis of sounds emitted in response to the same stimulus, some common characteristics emerge, as detailed in Sect. 4.2. For example, Fig. 3 shows

the spectrograms of samples that resulted to be the medioids of the three emitting context classes. Note that the dataset is imbalanced across contexts, due to the tendency of cats to not vocalize uniformly. Thus, such imbalances are indicative of their behavior during cat-human interactions.

Additional files can be found in the repository in a separate folder, with the following naming convention: `C_NNNNN_BB_SS_00000_FSEQX.wav`. These were the cases where the recorded cat meowed repeatedly, in sequences of many meows separated only by few seconds of silence.

4 Application Scenarios

CatMeows is open and available to the scientific community for research purposes. In this section, we present possible research directions starting from either our dataset or a similar collection of sounds, which can be built by following the protocol described in this paper.

Table 2. Number of files in the *CatMeows* dataset, grouped by situation, breed (MC = Maine Coon; EU = European Shorthair), and sex.

	Food (92)		Isolation (221)		Brushing (127)	
	MC (39)	EU (53)	MC (92)	EU (129)	MC (57)	EU (70)
Intact males (19)	–	5	10	–	4	–
Neutered males (76)	14	8	18	15	17	4
Intact females (68)	21	–	28	–	19	–
Neutered females (277)	4	40	36	114	17	66

Table 3. Number of files for each cat (F: waiting for food, I: isolation, B: brushing). A horizontal line separates Maine Coon and European Shorthair cats.

ID	F	I	B	Total
ANI01_MC_FN	–	4	6	10
BAC01_MC_MN	12	7	–	19
BRA01_MC_MN	2	10	7	19
BRI01_MC_FI	8	–	7	15
DAK01_MC_FN	4	32	4	40
JJX01_MC_FN	–	–	7	7
MEG01_MC_FI	4	10	–	14
NIG01_MC_MN	–	1	4	5
NUL01_MC_MI	–	10	4	14
WHO01_MC_FI	9	18	12	39

(continued)

Table 3. (*continued*)

ID	F	I	B	Total
BLE01_EU_FN	3	45	–	48
CAN01_EU_FN	2	26	26	54
CLE01_EU_FN	–	20	–	20
IND01_EU_FN	10	–	11	21
LEO01_EU_MI	5	–	–	5
MAG01_EU_FN	14	3	10	27
MAT01_EU_FN	6	10	10	26
MIN01_EU_FN	4	6	4	14
REG01_EU_FN	1	1	3	5
SPI01_EU_MN	8	15	4	27
TIG01_EU_FN	–	3	2	5

4.1 Proposed Scenarios

Among many application scenarios, the most obvious one is probably to support comparative and repeatable experimentation concerning our original project, that aimed at the automatic classification of cat vocalizations [17]. In detail, the idea is to have a fully automatic framework with the ability to process vocalizations and reveal the context in which they were produced, by using suitable audio signal processing techniques and pattern recognition algorithms.

Another usage of the *CatMeows* dataset may concern feature analysis in general. In this scenario, key features are detected in order to better understand what is meaningful in cat-human communication. Under this perspective, the original situations in which sounds have been recorded are only examples of contexts that cause cats to emit vocalizations. Moreover, a statistical analysis of the dataset’s audio features will be presented and discussed in Sect. 4.2.

Besides, since both cats and situations are univocally identified in the dataset, it is possible to plan listening sessions aiming to test the ability of humans in: i) recognizing specimens across different recordings, ii) clustering vocalizations emitted by different cats in the same context, iii) distinguishing between male/female cats or specimens belonging to different breeds, etc. In this sense, some early experiments have already been carried out [24].

Finally, a different scenario implies the investigation of meowing in relation with cat behavior. While our research took a given stimulus as the input and the corresponding vocalization as the output, deliberately ignoring behaviors and emotional states of cats, such aspects could be addressed in other research projects. To this end, even if in the current version of the dataset sounds are not tagged from the behavioral or emotional point of view, we are planning to make the corresponding video recordings available, thus facilitating in-depth affective analysis as well.

4.2 Example

To provide some insights on *CatMeows* and exemplify one of its possible uses, we carried out a statistical analysis of the dataset's audio features, and specifically: Vocalization duration D , Level Envelope E , Pitch F_0 , Tristimulus T_{1-3} , Roughness R , Brightness B , and Spectral Centroid C (for further information about audio features, please refer to [1]).

- D is just the duration of the vocalisation;
- E is the envelope profile of audio level, carrying information about dynamics (*i.e.* slow amplitude modulations);
- F_0 is the fundamental frequency of the sound. As we are dealing with meows, that are harmonic sounds, such a frequency is supposed to be always present;
- T_{1-3} is a feature borrowed from color perception and adapted to the audio domain. T_1 is the ratio between the energy of F_0 and the total energy, T_2 is the ratio between the second, third, and fourth harmonics taken together and the total energy, and T_3 is the ratio between remaining harmonics and the total energy [22]. Since the sampled signal is strongly band-limited, T_3 would have considered only few harmonics – if any – thus it was discarded. T_{1-2} was chosen to represent information regarding formants, since with a such band-limited signal, containing only a reduced number of harmonics, the actual computation of formants, either with filterbanks or linear prediction techniques, resulted to be reliable only in a reduced number of cases. The downside of using T_{1-3} is that this feature strongly depends on the quality of F_0 estimation;
- R is defined by [9] as the sensation for rapid amplitude variations, which reduces pleasantness, and whose effect is perceived as dissonant [21]. This is a timbral perceptual descriptor which may convey information about unpleasantness;
- C is the *center of mass* of the energy distribution across the spectrum. This feature provides relevant information about the spectral content, without considering harmonic aspects [12].

The pre-processing consisted in trimming the first and last 0.4 s of the audio file, and to perform a linear fade-in and fade-out of 0.05 s each. This was made in order to discard possible background sound events occurred during the 0.5 s handles before and after the cat vocalisation.

Except for D , all the above-mentioned features are time-varying signals, extracted either from the short-term Fourier Transform of the sound (for the frequency domain features), or an equal frame subdivision of the signal (for time-domain features), so as to have the same temporal resolution for every feature. In particular we used a window size of 512 samples, with an overlap of 448 samples (that is, an hop-size of 64 samples), and a Hanning windowing function.

- D is expressed in seconds, and computed as the length of the recording minus one (to remove the duration of silence handles);

- E is computed as the Root Mean Square (RMS) of each audio frame, and it is expressed in decibels;
- F_0 is computed with the SWIPE' method [4], limiting the results in the range 100 Hz–1250 Hz, and ignoring values where the confidence reported by SWIPE' is below 0.15. A median filtering with a span of 5 frames helped in removing artifacts. A visual inspection of the results showed that these settings works well, with only limited and sporadic octave-errors. F_0 is expressed in a logarithmic frequency scale;
- T_{1-2} are computed via retrieving the energy around F_0 and the corresponding harmonics using a triangular window of 9 FFT bins to mitigate potential F_0 errors and the FFT finite resolution. T_{1-2} are expressed in decibels;
- R is computed using the corresponding function of the MIRToolbox [14] and more specifically using the strategy proposed in [26].
- C is the weighted mean of the frequencies, using the energy over each frequency as the weight. C is expressed in a logarithmic frequency scale;

To avoid measuring features over silent portions of the signal which typically burden the modelling process, features that do not rely on F_0 (which implicitly ignores silence and non-periodic portions) were ignored in the instants where $E < 0.0005$ (−66 dB). This value has been empirically found to provide almost perfect silence detection over the entire dataset⁴.

In order to reduce the time-varying features to a set of global values, and to mitigate the effects of local errors (of any magnitude), median values and inter-quartile ranges are calculated for every feature. To grasp information about temporal aspects, this has been carrier out also to the delta-features (i.e. the differential of the feature in the temporal dimension). This led to a final feature space of a dimensionality equal to 25 variables.

Finally, a One-way ANOVA test has been carried out to assess which features are meaningful in distinguishing three different partitions of the dataset: Male v.s. Female; European Shorthair v.s. Maine Coon; Brushing v.s. Isolation v.s. Waiting for food. Results are reported in Table 4, where columns represent the comparison of two classes, rows represent the feature space, and a “•” mark points out where ANOVA revealed significant differences ($p < 0.05$).

From Table 2 it can be seen how isolation is the situation that elicits the majority of meows, and that European Shorthair cats tend to vocalize more than Maine Coon. Nevertheless, Table 3 shows a great variability among cats' reactions to given stimuli. Finally, Table 4 shows that every class can be distinguished: the emitting context affects all features; the breed mainly affects D , P , E , and C ; the sex mainly affects E , and C .

⁴ Please note that different envelope functions and frame sizes may have different optimal thresholds.

Table 4. ANOVA results, where “•” denotes cases with significant differences between classes. m subscript indicates the median; i subscript indicates interquartile range; Δ prefix indicates features’ delta. FI = Food-Isolation; FB = Food-Brushing; IB = Isolation-Brushing; EM = European Shorthair-Maine Coon; FM = Female-Male.

	FI	FB	IB	EM	FM
D	•	•		•	
P_m	•	•		•	
P_i			•	•	
ΔP_m			•	•	
ΔP_i	•	•	•	•	
E_m	•		•	•	
E_i				•	•
ΔE_m	•		•	•	•
ΔE_i	•		•	•	•
C_m				•	
C_i	•		•	•	•
ΔC_m	•		•	•	•
ΔC_i			•		
$T1_m$	•				
$T1_i$	•	•		•	
$\Delta T1_m$					
$\Delta T1_i$		•	•		
$T2_m$	•				
$T2_i$					
$\Delta T2_m$					
$\Delta T2_i$	•	•	•		
R_m	•		•		
R_i	•		•	•	•
ΔR_m					
ΔR_i	•		•		

5 Conclusions

This paper described the motivation, recording protocol, and contents of a dataset encompassing cat vocalizations in response to diverse stimuli. It comprises a fundamental step in cataloging and modeling cat-human communication, while, to the best of our knowledge, *CatMeows* is a unique corpus of audio documents.

After extensive analysis of the dataset’s contents, we provided potential application scenarios as well as an initial statistical analysis of physically/perceptually meaningful features which could address them. This does not limit the usage

of unsupervised feature-extraction mechanisms based on deep learning technologies [10].

Given the wide range of applications, *CatMeows* has not been divided in train-test sets beforehand; however, cross-validation techniques are recommended in order to evaluate the generalization capabilities of potential models. Focusing on reproducibility of the obtained results, which comprises a strong requirement in modern machine learning-based systems, the dataset is available to the scientific community. We hope that the availability of *CatMeows* will encourage further research in the scientific field of cat-human interaction and animal welfare in general.

Ethical Statement

The present project was approved by the *Organism for Animals Welfare* of the University of Milan (approval n. OPBA_25.2017). The challenging situations cats were exposed to were required in order to stimulate the emission of meows related to specific contexts. They were conceived considering potentially stressful situations that may occur in cats' life and to which cats can usually easily adapt. In order to minimize possible stress reactions, preliminary information on the normal husbandry practices (e.g., brushing or transportation) to which the experimental cats were submitted and on their normal reactions to these practices were collected by interviews to the owners. Additionally, cats were video recorded using a camera connected to a monitor for 5 min before the stimulus, during the stimulus and for 5 min after the stimulus, in order to monitor their behavior during the isolation challenge, with the idea of stopping the experiment if they showed signs of excessive stress; however, such a situation never occurred.

Intellectual Property of the Dataset

The dataset is one of the outcomes of an interdepartmental project at [OMITTED]. The dataset is open access, released under Creative Commons Attribution 4.0 International licence.⁵ It can be retrieved from the following DOI: [10.5281/zenodo.4007940](https://doi.org/10.5281/zenodo.4007940).

The dataset can be freely used for research and non-commercial purposes, provided that the authors are acknowledged by citing the present paper.

References

1. Alías, F., Socoró, J.C., Sevillano, X.: A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Appl. Sci.* **6**(5), 143 (2016)

⁵ <https://creativecommons.org/licenses/by/4.0/legalcode>.

2. Bradshaw, J., Cameron-Beaumont, C.: The signalling repertoire of the domestic cat and its undomesticated relatives. In: *The Domestic Cat: The Biology of Its Behaviour*, pp. 67–94 (2000)
3. Brown, S.L.: The social behaviour of neutered domestic cats (*Felis catus*). Ph.D. thesis, University of Southampton (1993)
4. Camacho, A., Harris, J.G.: A sawtooth waveform inspired pitch estimator for speech and music. *J. Acoust. Soc. Am.* **124**(3), 1638–1652 (2008)
5. Cameron-Beaumont, C.: Visual and tactile communication in the domestic cat (*Felis silvestris catus*) and undomesticated small-felids. Ph.D. thesis, University of Southampton (1997)
6. Car Working Group: Headset Profile (HSP) 1.2. Bluetooth SIG (2008)
7. Carney, H., Gourkow, N.: Impact of stress and distress on cat behaviour and body language. In: Ellis, S., Sparkes, A. (eds.) *The ISFM Guide to Feline Stress and Health*. Tisbury (Wiltshire): International Society of Feline Medicine (ISFM) (2016)
8. Casey, R.: Fear and stress in companion animals. In: Horwitz, D., Mills, D., Heath, S. (eds.) *BSAVA Manual of Canine and Feline Behavioural Medicine*, pp. 144–153. British Small Animal Veterinary Association, Guarantee (2002)
9. Daniel, P., Weber, R.: Psychoacoustical roughness: implementation of an optimized model. *Acta Acust. United Acust.* **83**(1), 113–123 (1997)
10. Dara, S., Tumma, P.: Feature extraction by using deep learning: a survey. In: *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1795–1801 (2018)
11. Eriksson, M., Keeling, L.J., Rehn, T.: Cats and owners interact more with each other after a longer duration of separation. *PLOS One* **12**(10), e0185599 (2017). <https://doi.org/10.1371/journal.pone.0185599>
12. Grey, J.M., Gordon, J.W.: Perceptual effects of spectral modifications on musical timbres. *J. Acoust. Soc. Am.* **63**(5), 1493–1500 (1978)
13. Karsh, E.B., Turner, D.C.: The human-cat relationship. *Domestic Cat: Biol. Behav.* 159–177 (1988)
14. Lartillot, O., Toivainen, P.: A matlab toolbox for musical feature extraction from audio. In: *International Conference on Digital Audio Effects*, pp. 237–244. Bordeaux (2007)
15. McComb, K., Taylor, A.M., Wilson, C., Charlton, B.D.: The cry embedded within the purr. *Curr. Biol.* **19**(13), R507–R508 (2009)
16. Notari, L.: Stress in veterinary behavioural medicine. In: *BSAVA Manual of Canine and Feline Behavioural Medicine*, pp. 136–145. BSAVA Library (2009)
17. Ntalampiras, S., et al.: Automatic classification of cat vocalizations emitted in different contexts. *Animals* **9**(8), 543.1–543.14 (2019). <https://doi.org/10.3390/ani9080543>
18. Owens, J.L., Olsen, M., Fontein, A., Kloth, C., Kershenbaum, A., Weller, S.: Visual classification of feral cat *Felis silvestris catus* vocalizations. *Curr. Zool.* **63**(3), 331–339 (2017)
19. Palestrini, C.: Situational sensitivities. In: Horwitz, D., Mills, D., Heath, S. (eds.) *BSAVA Manual of Canine and Feline Behavioural Medicine*, pp. 169–181. British Small Animal Veterinary Association, Guarantee (2009)
20. Palestrini, C., et al.: Stress level evaluation in a dog during animal-assisted therapy in pediatric surgery. *J. Veterinary Behav.* **17**, 44–49 (2017)
21. Plomp, R., Levelt, W.J.M.: Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38**(4), 548–560 (1965)

22. Pollard, H.F., Jansson, E.V.: A tristimulus method for the specification of musical timbre. *Acta Acust. United Acust.* **51**(3), 162–171 (1982)
23. Pongrácz, P.: Modeling evolutionary changes in information transfer: effects of domestication on the vocal communication of dogs (*Canis familiaris*). *Eur. Psychol.* **22**(4), 219–232 (2017)
24. Prato Previde, E., et al.: What’s in a meow? A study on human classification and interpretation of domestic cat vocalizations. *Animals* **10**, 1–17 (2020). in press
25. Telephony Working Group: Hands-Free Profile (HFP) 1.7.1, Bluetooth Profile Specification. Bluetooth SIG (2015)
26. Vassilakis, P.: Auditory roughness estimation of complex spectra-roughness degrees and dissonance ratings of harmonic intervals revisited. *J. Acoust. Soc. Am.* **110**(5), 2755 (2001)