Artificial Neural Network approach for revealing individuality, group membership and age information in goat kid contact calls

Published version:

DOI:10.3813/AAA.918758

(Article begins on next page)

ACTA ACUSTICA UNITED WITH ACUSTICA Vol. 100 (2014) 782 – 789

DOI:10.3813/AAA.918758

Artificial Neural Network approach for revealing individuality, group membership and age information in goat kid contact calls

Authors and affiliations

Livio Favaro[1], Elodie F. Briefer[2], Alan G. McElligott[3]

[1] Department of Life Sciences and Systems Biology, University of Torino, Via Accademia Albertina 13, 10123 Turin, Italy. livio.favaro@unito.it

[2] Institute of Agricultural Sciences, ETH Zürich, Universitätstrasse 2, 8092 Zürich, Switzerland. elodie.briefer@usys.ethz.ch

[3] Biological and Experimental Psychology, School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK. a.g.mcelligott@qmul.ac.uk

Corresponding authors: LF and AGM

Abstract

Machine learning techniques are becoming an important tool for studying animal vocal communication. The goat (*Capra hircus*) is a very social species, in which vocal communication and recognition are important. We tested the reliability of a Multi-Layer Perceptron (feed-forward Artificial Neural Network, ANN) to automate the process of classification of calls according to individual identity, group membership and maturation in this species. Vocalisations were obtained from 10 half-sibling (same father but different mothers) goat kids, belonging to 3 distinct social groups. We recorded 157 contact calls emitted during first week, and 164 additional calls recorded from the same individuals at 5 weeks. For each call, we measured 27 spectral and temporal acoustic parameters using a custom built program in Praat software. For each classification task we built *stratified* 10-fold cross-validated neural networks. The input nodes corresponded to the acoustic parameters measured on each signal. ANNs were trained with the error-back-propagation algorithm. The number of hidden units was set to the number of attributes + classes. Each model was trained for 300 epochs (learning rate 0.2; momentum 0.2). To estimate a reliable error for the models, we repeated 10-fold cross-validation iterations 10 times and calculated the average predictive performance. The accuracy was 71.13 ± 1.16% for vocal individuality, 79.59 ± 0.75% for social group and 91.37 ± 0.76% for age of the vocalising animal. Our results demonstrate that ANNs are a powerful tool for studying vocal cues to individuality, group membership and maturation in contact calls. The performances we achieved were higher than those obtained for the same classification tasks using classical statistical methods such as Discriminant Function Analysis. Further studies, investigating the reliability of these algorithms for the real-time classification of contact calls and comparing ANNs with other machine learning techniques are important to develop technology to remotely monitor the vocalisations of domestic livestock.

Keywords

bioacoustics, goats, livestock, machine learning, ungulates, vocal communication

PACS no. 43.80.Ka

1. Introduction

A crucial step in understanding animal vocal behaviour is the description and quantification of similarities and differences among acoustic signals [1, 2]. This step is essential in order to identify biologically meaningful categories of sound [1]. Indeed, several vocalisations encode a variety of information about animal sex, body size, age and even social status [3, 4, 5]. Vocal signal categorisation may also allow the detection of social context-dependent variability, ecological diversity, species recognition, and vocal individuality [6, 7]. Characterising animal sounds can also provide information on genetic and evolutionary relationships among different taxonomic units [8].

Traditionally, the classification of animal vocal signals has been performed using subjective methods [9], such as signal classification by multiple listeners using their pattern recognition abilities (e.g. the abilities of human observers to recognise vocal categories using their auditory system). More recently, technological improvements have allowed detailed acoustic measurements on recorded vocal parameters, followed by automated classification using statistical methods. These later, more advanced techniques, include Multivariate Analysis of Variance [10], Discriminant Function Analysis [11], Hierarchical Cluster Analysis [12], and Principal Components Analysis [13]. However, statistical methods frequently fail to detect biologically meaningful information in vocalisations [1]. A modern and alternative approach is to use mathematical computational techniques. Among these, machine learning (ML) algorithms have been suggested as an attractive, non-linear alternative to traditional statistical analyses [14, 15]. The biggest advantage of ML techniques is their ability to model complex and non-linear relationships among acoustic parameters without having to satisfy the restrictive assumptions required by conventional parametric approaches. Moreover, they allow modelling of non-linear associations with a variety of data types, and accommodate interactions among predictor variables with limited *a priori* specifications [16].

More recently, the reliability of ML techniques for solving complex pattern recognition problems has been demonstrated in many ecological [i.e. 17], biomedical [i.e. 18] and behavioural studies [i.e.

19]. Although the application of these approaches to bioacoustics has increased in the last decade, the growth has been slower than in other disciplines, and there is still a good degree of scepticism with respect to the role of these techniques in quantitative analyses [20].

Artificial Neural Networks (ANNs) are the most common ML methods used for classification and recognition of mammal vocalisations. These algorithms were firstly introduced in marine bioacoustics to study the sonar system of bottlenose dolphins, *Tursiops truncatus* [21], and were further used for recognition of vocal units, caller and species in many different marine mammals [1, 22, 23, 24]. ANNs have been successfully used to identify echolocating bat species [25], to classify several non-human primates vocalisations [26, 27], Gunnison's prairie dog, *Cynomys gunnisoni* [28], fallow deer, *Dama dama* [14], and even stress-linked calls of domestic pigs, *Sus scrofa* [29].

Domestic goats are very social animals, and vocal communication and recognition are important for social bonding and group cohesion [5, 30]. Goat kids produce one basic call type, the “contact call”, when isolated at short distance from other group members [5]. According to the source-filter theory of voice production [31, 32], calls are generated by vibrations of the vocal folds (source, determining the fundamental frequency, “F0”) and are subsequently filtered by the supralaryngeal vocal tract (filter, resulting in amplified frequencies called “formants”). In goat kid vocalisations, F0 and formants parameters are known to be good indicators of caller identity [33], group membership [30], body size, sex and even age [5].

In this study, we tested the reliability of a Multi-Layer Perceptron (MLP), feed forward ANN, to automate classification of calls according to individual identity, group membership and maturation in a livestock species: the goat (*Capra hircus*). Our aim was to determine whether the MLP performances were better than those obtained for the same classification tasks using classical statistical methods such as Discriminant Function Analysis and, therefore, should be adopted in future vocal communication studies.

2. Methods

2.1 *Animals and recordings*

Contact calls were collected from 10 goat kids (9 males and 1 female), belonging to 3 distinct social groups at White Post Farm, Nottinghamshire, UK (53°06'N, 1°03'W). Goats use both open and closed mouth contact calls but, for the purpose of this study, we used only open mouthed calls, since closed mouth calls suppress or modify some formants [5]. All kids were half-sibling (same father but different mothers) born in July and December 2009, and March 2010, respectively. Each group (6.00 +/- 0.97 kids per group, mean +/- SE) was housed in an indoor communal pen of 4.4 m x 4.5 m. Vocalisations were recorded from the same individuals both early after parturition (1 week ± 5 days), and when young kids (5 weeks ± 5 days). To promote contact call production, we isolated kids from their mothers for 5 min periods, 2-3 times per day between 10 and 5 pm. The distance to the mother was set at 1 m (on average), during the first day of recordings and increased afterwards if necessary, until we obtained contact calls [i.e. low-affect vocalizations, 5] instead of distress calls [i.e. high-pitched vocalizations associated with high stress levels, 5]. Kids were isolated alone, except if they showed signs of stress during isolation even at 1 m. In these cases, they were isolated with their sibling(s).

Recordings were collected with a Sennheiser MKH70 directional microphone (frequency response 50 Hz to 20 kHz ± 2.5 dB) connected to a Marantz PMD660 digital recorder (sampling rate set to 44.1 kHz). During recording sessions, the microphone was placed at distances of 1 - 5 m from the vocalising animal. Segments containing acoustic recordings were saved in WAV format (16-bit amplitude resolution) and stored into an SD memory card. All the files were then transferred to a computer for later acoustic analyses.

2.2 *Acoustic analysis*

For each file, the waveform and FFT spectrogram (window length = 0.01 s, time steps = 1000, frequency steps = 250, Gaussian window shape, dynamic range = 50 dB) were generated in Seewave [34]. After visual examinations of sonograms, calls with high background noise levels were discarded. Among the remaining vocalizations, we selected 157 good quality contact calls (13 to 23 calls per individual) emitted during early postnatal days, and 164 additional calls (13 to 24 calls per individual) recorded from the same individuals at 5 weeks (Table 1, Figure 1).

For each call, we measured 27 spectral and temporal acoustic parameters (Table 2), which were potentially important for vocal distinctiveness. These included both temporal measures, such as call duration, related to lung capacity [35], source-related vocal features (F0) and filter-related acoustic vocal features (formants), [5, 32]. Acoustic measurements were carried out using a custom built program [36, 37] in Praat v.5.0.47 DSP Package [38].

<u>Source-related parameters</u>

We extracted the F0 contour of each call using a cross-correlation method ([Sound: To Pitch (cc) command], 1 week old: time step = 0.005 s, pitch floor =300-400 Hz, pitch ceiling = 700-900 Hz; 5 weeks old: time step = 0.005-0.015 s, pitch floor =200-300 Hz, pitch ceiling = 700-800 Hz). If the entire F0 contour could not be detected, calls were high-pass filtered before the analysis (cut-off frequency: 1 week old, 300 Hz; 5 weeks old, 200 Hz). For each extracted F0 contour, we measured the following vocal parameters: the frequency value of F0 at the start (F0Start) and at the end (F0End) of the call; the mean (F0Mean), minimum (F0Min) and maximum (F0Max) F0 frequency values across the call; the percentage of the total call duration when F0 was maximum (TimeF0Max); and the F0 mean absolute slope (F0AbsSlope). Moreover, we calculated F0 variation by measuring jitter (the mean absolute difference between frequencies of consecutive F0 periods divided by the mean frequency of F0 [Jitter (local) command]) and shimmer (the mean absolute difference between the amplitudes of consecutive F0 periods divided by the mean amplitude of F0 [Shimmer (local) command]) parameters.

<u>Filter-related parameters</u>

We extracted the contour of the first four formants of each call using Linear Predictive Coding analysis (LPC; [Sound: To Formant (burg) command], 1 week old: time step = 0.003 s, maximum number of formants = 4–5, maximum formant = 9800–12000 Hz, window length = 0.01–0.04 s; 5 weeks old: time step = 0.01-0.025 s, maximum number of formants = 4–5, maximum formant = 8000–10000 Hz, window length = 0.01–0.05 s). To check if the Praat software accurately tracked the formants, the outputs of the LPC analysis were visually inspected together with the spectrograms. Spurious values and inter-segment values were deleted and we corrected for octave jumps when necessary. For each call we collected the mean (F1-4Mean) minimum (F1-4Min), and maximum (F1-4Max) values of the formants. Further, we estimated the minimum formant dispersion (DfMin) and the vocal tract length of vocalising kids (estVTL) using the methods described by Reby and McComb [36] and validated for goats by Briefer & McElligott [5]. Finally, we measured the frequency values at the upper limit of the first (Q25%), second (Q50%) and third (Q75%) quartiles of energy, using a linear amplitude spectrum applied to the whole call, and we included in the analyses the total duration of each call (Dur).

2.3 *Classification tasks*

We tested the reliability of a neural network to automate classification of goat kids contact calls according to:

1) Caller individual identity
2) Caller group membership
3) Caller age

For the classification tasks 1 and 2, we used 157 contact calls recorded when goat kids were 1 week old (Table 1). For the classification task 3, we introduced in the analysis 164 calls recorded from the same individuals at 5 weeks of age (Table 1).

2.4 *Artificial Neural Network*

Architecture

For this study, we used a supervised Multi-Layer Perceptron (MLP), feed-forward artificial neural network, computed in the WEKA v. 3.6.9 software package [39]. The MLP was trained with the error-back-propagation method developed by Rumelhart et al. [40]. In this MPL architecture, the processing elements are arranged in the following layered structure: (a) the input nodes, (b) the hidden layers and (c) an output layer. Each neuron is connected to the other adjacent elements by axons, and the signals are transmitted forward only: from the input nodes to the output neurons through the hidden layers. The input nodes of our MLP corresponded to the acoustic parameters measured on each contact call. The output neurons corresponded, for the three classification tasks, to the identities of the callers, the group memberships of the callers, and the ages of the callers, respectively. A schematic representation of the Multi-Layer Perceptron used is presented in Figure 2.

Training and testing

For each of the three classification tasks, the number of hidden units was set to the number of attributes + classes. Each model was trained for 350 epochs (learning rate 0.2; momentum 0.2). We determined these optimal values empirically, by studying the performances of different cross-validated MLP with a trial-and-error approach [26, 14]. We used 10-fold cross-validation to build robust models. For each classification task, the dataset was randomly reordered and then split into 10 folds of equal size. In each iteration, one fold was used for the testing phase and the other 9 folds for the training phase. In particular, we performed a *stratified* cross-validation. This means that folds were created to reflect the same class distributions in each fold as in the complete dataset. We chose this approach, because non cross-validated machine-learning algorithms are likely to overfit the training, and to lose their accuracy and ability to generalize during the test phase [41]. To estimate a reliable error of the models, we repeated 10-fold *stratified* cross-validation iterations 10 times and calculated the average predictive performance. Finally, before building the models, all the features were scaled by applying the feature normalization algorithm implemented in WEKA. This pre-processing procedure can improve ANN efficiency by keeping the connection weights from becoming too large and swamping the model during training phase [42].

ANN performance evaluation

The performances of the model were assessed (for each classification task) by calculating the following three retrieval metrics:

1) Accuracy (ACC). This value shows the percentage of test instances that were correctly classified by the neural network;

2) Kappa statistic (kappa). This value assesses whether the performance of the neural network differed from expectations based on chance alone [43, 44]. Kappa can vary between 1 (perfect classification) and 0 (classification expected by chance);

3) Area under the receiver operating characteristic curve (AUC). The AUC of a classifier is equivalent to the probability that it ranks a randomly chosen positive instance higher than a randomly chosen negative one. AUC values can vary between 0 and 1.

It is important to note that, since kappa and AUC metrics are computed in WEKA for the binary class case, we handled the multiclass classification (caller identity and group membership tasks) using the “one against all" strategy. In particular, we treated each class value in turn as the "positive" class and all others as the "negative" class.

Comparison of ANN with Discriminant Function Analysis (DFA)

In order to provide a direct comparison of the ANN accuracy with a more classical multivariate technique, we performed a discriminant function analysis in SPSS v. 19 (SPSS, Inc. 2010) for each of the three classification tasks, using the same dataset presented to the MLP. Firstly, to meet the assumption of independence between predictor variables, we performed a principal component analysis (PCA). Principal Components (PC) showing eigenvalues > 1 were used to classify vocalisations with a cross-validated (leave-one-out) DFA.

3. Results

The Multi-Layer Perceptron succeeded in classifying most of the contact calls according to individuality, group membership and age of the goat kids. The average and standard deviations of the ACC showed limited variation within each classification task (Table 3). Average predictive performances for each classification task were, respectively, 71.13 ± 1.16 % for the caller individual identity (N = 10 individuals and 157 calls), 79.52 ± 0.76 % for the caller group membership (N = 10 individuals and 157 calls) and 91.37 ± 0.76 % for the caller age (N = 10 individuals and 321 calls). The average kappa and AUC values of the neural network models were, respectively: caller identity task = 0.62 ± 0.02 and 0.78 ± 0.03, group membership = 0.68 ± 0.01 and 0.92 ± 0.01, caller age = 0.85 ± 0.02 and 0.98 ± 0.01 (Figure 3). The PCA explained 78.33 % of the total variability with 5 PCs showing eigenvalues exceeding 1. The cross-validated DFA performed using this PCA factor solution correctly classified 43.0 % of the vocal signals according to the caller individual identity, 73.50 % according to the caller group membership and 87.50 % to the caller age. To summarise, the Multi-Layer Perceptron used in this study achieved a higher accuracy than the DFA and yielded reliable predictions (none based on chance), in classifying the contact calls according to individuality, group membership and age of emitters.

4. Discussion

We investigated whether a Multi-Layer Perceptron (MLP), feed-forward artificial neural network (ANN), could reliably classify goat kids contact calls according to the caller identity, group membership and age. To this end, we used a database of vocalisations recorded from 10 kids during the immediate postnatal period (1 week) and additional calls recorded from the same individuals at 5 weeks. For each vocalisation, we measured 27 spectral and temporal acoustic parameters, which were then presented to the neural network as input variables. The MLP showed a higher level of accuracy (ACC) compared to the results obtained with the cross-validated DFA. In particular, the DFA correctly classified 43.0 % of 1 week kid calls according to the emitter, while the MLP achieved an average ACC of 71.13 %. The MLP obtained a higher ACC also in the group membership identification of the caller (79.52 % vs 73.50 %) and suggested the presence of a social effect on the ontogeny of vocalisations in this species. Accordingly, Briefer and McElligott [30] showed that the social group influenced the energy distribution in the spectrum (energy quartiles) and the second and third formants. This probably results from changing the shape and length of the vocal tract. Finally, the MLP proved more reliable than the DFA also in classifying calls according to caller age (91.37 % vs 87.50 %), revealing the age-related changes in the vocal parameters of contact calls [5]. Overall, the MLP accuracy performances suggest that these algorithms can be used as a modern and reliable alternative to traditional statistical methods in bioacoustics.

The MLP we used showed average kappa values of 0.62 ± 0.02 (caller individuality task), 0.68 ± 0.01 (group membership task), 0.85 ± 0.02 (caller age task). Fleiss [44] suggested that kappa values greater than 0.75 can be considered to represent excellent agreement beyond chance, values below 0.40 indicate poor agreement beyond chance, and values between 0.40 and 0.75 may be taken to represent fair to good agreement beyond chance. According to Fleiss [44], we suggest that the MLP presented in this study show reliable predictions and matching not based on chance, in each of the three classification tasks. The average area under the receiver operating characteristic curve (AUC) values were 0.78 ± 0.03 (caller individuality task), 0.92 ± 0.01 (group membership task) and 0.98 ± 0.01 (caller age task) respectively. Fawcett [45] showed that random guessing classification produces an AUC of 0.5 and suggested that realistic classifiers should have AUC higher than 0.5. Accordingly, we consider the values observed in this study as a very good discrimination in each classification task.

Overall, our results confirmed that the MLP can process a variety of spectral and temporal acoustic parameters to classify vocal signals [26]. In particular, we used temporal measures related to lung capacity (i.e. duration), source-related vocal features (F0) and filter-related acoustic vocal features (formants) to show that the MLP can be used to study vocalisations from a source-filter perspective [35]. Moreover, although ANNs have been previously used in the study of wild mammal vocalisations [27], very few reports exist for the use and potential of these techniques in farm animal research [29]. In particular, our study is the first to show the reliability of these algorithms for the classification of domestic livestock vocalisations. Developing novel tools to understand which animals are calling and to extract biological meaningful information from vocalisations has great potential for remotely monitoring domestic livestock, especially on farms with large numbers of animals. In future, the technology could be used to investigate whether the calls uttered indicate that the animals are in positive or negative states, and even to investigate their emotions [46].

We used a MLP to analyse a particular animal call type: the contact call. Contact calls are very complex signals, mostly used by birds and mammals, encoding a great deal of information about the emitter [47]. The results achieved by the MLP in grouping these calls provide evidence that ANN algorithms have the capacity to extract and categorise the biological meaningful information encoded in mammal vocalisations.

In conclusion, our results show successful examples of signal recognition by a MLP for individuality, group membership and maturation in domestic goat kids, suggesting that ANNs could be considered a reliable tool to study vocalisations of domestic livestock from a source-filter perspective. ANNs also have the potential to exhibit substantially greater predictive power than traditional statistical approaches and we argue that these algorithms can be adopted to classify contact calls of many different species. Further research, comparing ANNs with other machine learning techniques would be especially valuable. We also recommend additional investigations to evaluate whether ANNs could classify contact calls in real-time and therefore be suitable to develop effective passive acoustic monitoring systems.

5. Acknowledgments

The authors would like to thank E. Antill, C. Booth, E. Cant, C. Charpin, K. Cho Geun-A, C. Farrington, F. Galbraith, E. Landy, M. Padilla de la Torre and M. Wang for their help with data collection. We are grateful to D. Reby for providing the custom built program in Praat. Special thanks are due to D. Stowell and D. Pessani for their helpful comments. We thank the staff of White Post Farm (http://whitepostfarmcentre.co.uk/) for their help and free access to their animals. L. Favaro was supported by the University of Torino through a Fiat Group Automobiles S.p.A. research grant. E. Briefer was funded by a Swiss National Science Foundation fellowship during data collection. We acknowledge the financial support of the University of London Central Research Fund for recording equipment.

6. References

[1] V. B. Deecke, V. M. Janik: Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls. J. Acoust. Soc. Am. **119** (2006) 645-653,.

[2] M. V. Torriani, E. Vannoni, A. G. McElligott: Mother-young recognition in an ungulate hider species: a unidirectional process. Am. Nat. **168** (2006) 412-420.

[3] M. B. Manser, R. M. Seyfarth and D. L. Cheney: Suricate alarm calls signal predator class and urgency. Trends Cogn. Sci. **6** (2002) 55-57.

[4] E. Briefer, E. Vannoni, A. G. McElligott. Quality prevails over identity in the sexually selected vocalisations of an ageing mammal. BMC Biology **8** (2010) 35.

[5] E. Briefer, A. G. McElligott: Indicators of age, body size and sex in goat kid calls revealed using the source-filter theory. Appl. Anim. Behav. Sci. **133** (2011) 175-185.

[6] V. M. Janik, L. S. Sayigh, R. S. Wells: Signature whistle shape conveys identity information to bottlenose dolphins. Proc. Natl. Acad. Sci. USA, **103** (2006) 8293-8297.

[7] E. Vannoni, A. G. McElligott: Individual acoustic variation in fallow deer (*Dama dama*) common and harsh groans: a source-filter perspective. Ethology 1**13** (2007) 223–234.

[8] J. B. Stachowicz, E. Vannoni, B. J. Pitcher, E. F. Briefer, E. Geffen, A. G. McElligott: Acoustic divergence in the rut vocalisation of Persian and European fallow deer. J. Zool. **292** (2014) 1-9.

[9] L. S. Sayigh, P. L. Tyack, R. S. Wells, M. D. Scott: Signature whistles of free-ranging bottlenose dolphins *Tursiops truncatus*: Stability and mother-offspring comparisons. Behav. Ecol. Sociobiol. **26** (1990) 247-260.

[10] F. Range, J. Fischer: Vocal repertoire of Sooty Mangabeys (*Cercocebus torquatus atys*) in the Tai National Park. Ethology **110** (2004) 301-321.

[11] E. C. Garland, A. W. Goldizen, M. L. Rekdahl, R. Constantine, C. Garrigue, N. Daeschler Hauser, M.M. Poole, J. Robbins, M. J. Noad: Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. Curr. Biol. **21** (2011) 687-691.

[12] V. M. Janik: Pitfalls in the categorization of behaviour: A comparison of dolphin whistle classification methods. Anim. Behav. **57** (1999) 133-143.

[13] J. J. Gros-Louis, S. E. Perry, C. Fichtel, E. Wikberg, H. Gilkenson, S. Wofsy, A. Fuentes: Vocal repertoire of *Cebus capucinus*: acoustic structure, context, and usage. Int. J. Primatol. **29** (2008) 641-670.

[14] D. Reby, S. Lek, I. Dimopoulos, J. Joachim, J. Lauga, S. Aulagnier: Artificial neural networks as a classification method in the behavioural sciences. Behav. Process. **40** (1997) 35-43.

[15] M. A. Acevedo, C. J. Corrada-Bravo, H. Corrada-Bravo, L. J. Villanueva-Rivera, T. M. Aide: Automated classi!cation of bird and amphibian calls using machine learning: A comparison of methods. Ecol. Inform. **4** (2009) 206-214.

[16] J. D. Olden, J. J. Lawler, N. L. Poff: Machine learning methods without tears: a primer for ecologists. Q. Rev. Biol. **83** (2008) 171–193.

[17] T. Tirelli, L. Favaro, M. Gamba, D. Pessani: Performance comparison among multivariate and data mining approaches to model presence/absence of *Austropotamobius pallipes* complex in Piedmont (North Western Italy). C. R. Biol. **334** (2011) 695-704.

[18] M. Khanmohammadi, A. B. Garmarudi, K. Ghasemi: Back-propagation artificial neural network and attenuated total reflectance-Fourier transform infrared spectroscopy for diagnosis of basal cell carcinoma by blood sample analysis. J. Chemom. **23** (2009) 538-544.

[19] S. Ghirlanda, M. Enquist: Artificial neural networks as models of stimulus control. Anim. Behav. **56** (1998) 1383–1389.

[20] S. Huebner: Bioacoustic Classifier System Design as a Knowledge Engineering Problem. In: Computational Bioacoustics for Assessing Biodiversity. K.H. Frommolt, R. Bardeli, M. Clausen (Eds.). Federal Agency for Nature Conservation, Vilm, Germany, 2008.

[21] W. W. Au: Comparison of sonar discrimination: dolphin and an artificial neural network. J. Acoust. Soc. Am. **95** (1994) 2728-2735.

[22] E. Mercado III, A. Kuh: Classification of humpback whale vocalizations using a self-organizing neural network. Proceedings of the International Joint Conference on Neural Networks **2** (1998) 1584-1589.

[23] S. O. Murray, E. Mercado, H.L. Roitblat: The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations. J. Acoust. Soc. Am. **104** (1998) 3626–3633.

[24] M. Marcoux, M. Auger-Méthé, M. M. Humphries: Variability and context specificity of narwhal (*Monodon monoceros*) whistles and pulsed calls. Mar. Mammal Sci. **28** (2011) 649-665.

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

[26] L. Pozzi, M. Gamba, C. Giacoma: The use of Artificial Neural Networks to classify primate vocalizations: a pilot study on black lemurs. Am. J. Primatol. **72** (2010) 337-348.

[27] A. Mielke, K. Zuberbühler: A method for automated individual, species and call type recognition in free-ranging animals. Anim. Behav. **86** (2013) 475-482.

[28] J. Placer, C. N. Slobodchikoff: A fuzzy-neural system for identification of species-specific alarm calls of Gunnison's prairie dogs. Behav. Process. **52** (2000) 1-9.

[29] P. C. Schön, B. Puppe, G. Manteuffel: Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*). J. Acoust. Soc. Am. **110** (2001) 1425-1431.

[30] E. F. Briefer, A. G. McElligott: Social effects on vocal ontogeny in an ungulate, the goat, *Capra hircus*. Anim. Behav. **83** (2012) 991-1000.

[31] G. Fant: Acoustic theory of speech production. The Hague, Mouton, 1960.

[32] A. Taylor, D. Reby: The contribution of source-filter theory to mammal vocal communication research. J. Zool. **280** (2010) 221-236.

[33] E. Briefer, A. G. McElligott: Mutual mother-offspring vocal recognition in an ungulate hider species (*Capra hircus*). Anim. Cogn. **14** (2011) 585-598.

[34] J. Sueur, T. Aubin, C. Simonis: Equipment Review: Seewave, a free modular tool for sound analysis and synthesis. Bioacoustics **18** (2008) 213-226.

[35] W. T. Fitch: Production of Vocalizations in Mammals. *In*: Encyclopedia of Language and Linguistics: 115-121, Brown K. (Ed.), Oxford, Elsevier, 2006.

[36] D. Reby, K. McComb: Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. Anim. Behav. **65** (2003) 519-530.

[37] B. D. Charlton, Z. Zhihe, R. J. Snyder: Vocal cues to identity and relatedness in giant pandas (*Ailuropoda melanoleuca*). J. Acoust. Soc. Am. **126** (2009) 2721-2732.

[38] P. Boersma: Praat, a system for doing phonetics by computer. Glot International **5** (2001) 341-345.

[39] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten: The WEKA Data Mining Software: An Update. SIGKDD Explorations **11** (2009) 1-18.

[40] D. E. Rumelhart, G. E. Hinton, R. J. Williams: Learning representations by back-propagating errors. Nature **323** (1986) 533-536.

[41] C. Schittenkopf, G. Deco, W. Brauer: Two Strategies to Avoid Overfitting in Feedforward Networks. Neural Networks **10** (1997) 505-516.

[42] R. C. Eberhart, R. W. Dobbins: Neural Network PC Tools. Academic Press, New York, NY, 1990.

[43] K. Titus, J. A. Mosher, B.K. Williams: Chance corrected classification for use in discriminant analysis. Am. Midl. Nat. **111** (1984) 1-7.

[44] J. L. Fleiss: Statistical methods for rates and proportions. Second Edition. John Wiley, New York, 1981.

[45] T. Fawcett: An introduction to ROC analysis. Pattern Recogn. Lett. **27** (2006) 861–874.

[46] E. Briefer: Vocal expression of emotions in mammals: mechanisms of production and evidence. J. Zool. **288** (2012) 1–20.

[47] N. Kondo, S. Watanabe: Contact calls: information and social function. Jpn. Psychol. Res. **51** (2009) 197–208.

FIGURE CAPTIONS

Figure 1. Spectrogram (window length: 0.009 s, time steps = 1000, frequency steps = 500, Gaussian window shape, dynamic range = 70 dB) of two contact calls recorded from the same goat kid at 1 week (left) and 5 weeks (right) of age. F0 indicates the fundamental frequency while F1-F4 indicate formant frequencies.

Figure 2. Schematic representation of the feed-forward Multi-Layer Perceptron.

Figure 3. Kappa statistic (kappa) and area under the ROC curve (AUC) values obtained for each classification task (I = Caller individual identity, M = Caller group membership, A = Caller age). T-bars represent 95% confidence interval.

TABLES

Table 1. Group membership and number of calls recorded for each goat kid at 1 week (*N* = 157) and 5 weeks of age (*N* = 164).

Goat kid	Group membership	1 week	5 weeks
1	A	15	16
2	A	15	15
3	A	15	16
4	B	13	13
5	B	15	16
6	B	14	15
7	C	23	24
8	C	15	15
9	C	17	16
10	C	15	18

Table 2. Abbreviations and brief descriptions of the vocal parameters measured on each call.

Abbreviation	Parameter
F0Start (Hz)	Frequency value of F0 at the start of the call
F0End (Hz)	Frequency value of F0 at the end of the call
F0Mean (Hz)	Mean F0 frequency value across the call
F0Min (Hz)	Minimum F0 frequency value across the call
F0Max (Hz)	Maximum F0 frequency value across the call
%TimeF0Max (%)	Percentage of the total call duration when F0 is maximum
F0AbsSlope (Hz/s)	F0 mean absolute slope
Jitter (%)	Mean absolute difference between frequencies of consecutive F0 periods divided by the mean frequency of F0
Shimmer (%)	Mean absolute difference between the amplitudes of consecutive F0 periods divided by the mean amplitude of F0
F1Mean (Hz)	Mean frequency value of the first formant
F2Mean (Hz)	Mean frequency value of the second formant
F3Mean (Hz)	Mean frequency value of the third formant
F4Mean (Hz)	Mean frequency value of the fourth formant
F1Min (Hz)	Minimum frequency value of the first formant
F2Min (Hz)	Minimum frequency value of the second formant
F3Min (Hz)	Minimum frequency value of the third formant
F4Min (Hz)	Minimum frequency value of the fourth formant
F1Max (Hz)	Maximum frequency value of the first formant
F2Max (Hz)	Maximum frequency value of the second formant
F3Max (Hz)	Maximum frequency value of the third formant
F4Max (Hz)	Maximum frequency value of the fourth formant
DfMin (Hz)	Minimum spacing of the formants
EstVTL	Estimation of the vocal tract length
Q25% (Hz)	Frequency value at the upper limit of the first quartiles of energy
Q50% (Hz)	Frequency value at the upper limit of the second quartiles of energy
Q75% (Hz)	Frequency value at the upper limit of the third quartiles of energy
Dur (s)	Duration of the call

Table 3. Accuracy (ACC) for each classification task. Each RUN is an average among 10 different *stratified* cross-validations.

	Caller individual identity	**Caller group membership**	**Caller age**
RUN	ACC	ACC	ACC
1	73.46	79.22	91.29
2	69.70	80.98	90.97
3	71.00	80.10	91.59
4	70.12	78.74	90.35
5	70.37	78.95	91.58
6	72.16	78.77	90.64
7	71.01	79.19	90.66
8	72.13	80.10	91.57
9	71.23	80.08	92.83
10	70.12	79.02	92.21
Average	71.13	79.52	91.37
St. Dev.	1.16	0.75	0.76