Animal Behaviour 86 (2013) 475–482

Contents lists available at SciVerse ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Commentary

A method for automated individual, species and call type recognition in free-ranging animals

Alexander Mielke[a,b,*], Klaus Zuberbühler[a,c]

[a] School of Psychology & Neuroscience, University of St Andrews, St Andrews, U.K.
[b] Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[c] Department of Comparative Cognition, Institute of Biology, University of Neuchatel, Neuchatel, Switzerland

ARTICLE INFO

Article history:
Received 4 December 2012
Initial acceptance 11 January 2013
Final acceptance 8 April 2013
Available online 24 June 2013
MS. number: 12-00913R

Keywords:
artificial neural network
blue monkey
Cercopithecus mitis
mel frequency cepstral coefficient
voice recognition

The ability to identify individuals reliably is often a key prerequisite for animal behaviour studies in the wild. In primates, recognition of other group members can be based on individual differences in the voice, but these cues are typically too subtle for human observers. We applied a combined mechanism consisting of a call feature extraction (mel frequency cepstral coefficients) and pattern recognition algorithm (artificial neural networks) to investigate whether automated caller identification is possible in free-ranging primates. The mechanism was tested for its accuracy in recognizing species, call type and caller identity in a large population of free-ranging blue monkeys, *Cercopithecus mitis stuhlmanni*, in Budongo Forest, Uganda. Classification was highly accurate with 96% at the species, 98% at the call type and 73% at the caller level. It also outperformed conventional discriminant function analysis in the individual recognition task. We conclude that software based on this method will make a powerful tool for future animal behaviour research, as it allows for automatic, fast and objective classifications in different animal species.

For many questions in animal behaviour, individual recognition of subjects is essential. To this end, fieldworkers usually rely on individual differences in body shape, coloration and markings. Although voice and chemical cues are often also individually distinct, they are more difficult to observe directly (Tibbetts & Dale 2007). In most field studies, researchers thus go through a time- and resource-consuming learning process or need to rely on artificial markings, such as rings and radiotracking, which can be difficult to administer and are almost always invasive for the animal (Adi et al. 2010). While there has been a strong focus on identifying individuals visually, vocalizations have great potential to facilitate recognition, especially for species that are cryptic, arboreal or nocturnal (Bardeli et al. 2010).

In primates and other groups of animals that live in stable social groups, individual recognition is a well-documented, key aspect of social behaviour (Tibbetts & Dale 2007). Moreover, in species living in environments with restricted visual contact, individual recognition is often based on vocal communication. Accordingly, individual vocal signatures have been reported in a number of primate species (e.g. *Macaca mulatta*: Hammerschmidt et al. 2000; *Papio ursinus*: Fischer et al. 2001; *Pan troglodytes*: Kojima et al. 2003; *Macaca fuscata*: Ceugniet & Izumi 2004). Playback studies have further shown that primates actively employ this information during social interactions (Lemasson et al. 2005; Seyfarth & Cheney 2008).

Despite this strong evidence for widespread vocal recognition in primates, as well as other animals, decisive steps have not yet been taken to develop software that can reliably replicate this ability, with its obvious benefits for fieldwork. For example, voice-based individual recognition would enable researchers to bypass the time-consuming process of learning to distinguish individuals and to carry out online identification of individuals that are not directly visible. Other potential benefits of automated caller identification are for research projects that involve large amounts of audio recordings, which are extremely time consuming to analyse. Here, individuals could be tracked using passive audio-recording equipment, which would be useful in helping to estimate home range size and use. New research questions could be asked while working with nonhabituated groups, such as estimating the length of tenure for males and other important demographic variables (Butynski et al. 1992). Finally, in habitats occupied by closely related species, automated caller recognition procedures may help in recognizing different species, which has census applications.

* Correspondence: A. Mielke, Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany.
E-mail address: Alexander_Mielke@eva.mpg.de (A. Mielke).

0003-3472/$38.00 © 2013 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.anbehav.2013.04.017

In animal vocalization studies, information about a call is often extracted manually from its spectrogram, while the choice of parameters is often driven by the intuition of the researcher, potentially eliminating valuable information. The manual extraction makes the process unsuitable for online identification and analysis of large data sets. Call classification is usually executed using discriminant function analysis (DFA), based on the assumption that the extracted features are separable into subsets through the linear division of class patterns along linear planes in space (Bortz 2005). Possible nonlinear applications of DFA (Mika et al. 1999) have not yet been used in animal call recognition research. As the perception and classification by conspecifics might be nonlinear, the DFA might not be optimal to model recognition processes (Deecke & Janik 2006).

In this paper, we introduce a method for automated species, call type and caller identity identification consisting of a feature extraction mechanism and a classifier. Feature extraction is the process of transforming each call from a high-dimensional to a low-dimensional vector (Bahoura & Simard 2010), preserving enough information for further processing, while the classifier algorithm operates to assign each call to one of several predefined categories. In developing the method, we took advantage of recent developments in human speaker and speech recognition, which is based on instantaneous processing of speech samples, followed by a robust classification process (Anusuya & Katti 2009). A key attribute is the fast, automated and standardized extraction of features from a sound signal, to reduce the amount of information encoded to a low-dimensional feature space (Campbell 1997). The most widely used method in current human speech and speaker recognition systems involves mel frequency cepstral coefficients (MFCC; Beigi 2011). MFCCs, instead of focusing on certain spectral features, map the entire spectrum by slicing it along the time and frequency axes and assigning values to the resulting cells based on the amplitude of the signal in that cell. Various studies have shown that MFCCs can be employed to classify animal signals (African elephants, *Loxodonta africanus*: Clemins & Johnson 2003; Clemins et al. 2005; birds: Kogan & Margoliash 1998). In contrast to classification approaches based on manually extracted spectral features, the extraction process is fully automated, repeatable and standardized, which makes it particularly attractive for field applications (Cheng et al. 2010). Additionally, the low number of a priori assumptions about the features makes it possible to apply the same algorithm for different call types and species.

The second key component consists of a classification algorithm that is able to operate on the MFCC output. Here, we opted for an artificial neural networks (ANN) approach. ANNs are able to learn associatively, generalize and recognize patterns by using simple units ('neurons') with weighted connections, which enables them to respond to information in differentiated ways (Ghirlanda & Enquist 2007). Like other classifiers, ANNs extract a general pattern from a training set of vocalizations for which category membership is known, which is then used to classify unknown calls. In contrast to DFAs, ANNs do not make any assumption about underlying probability distributions of the input vector and they can map input nonlinearly (Reby et al. 1997). ANN-based algorithms have been used in combination with different feature extraction methods for recognition of callers, call types and species in marine mammals (Mercado & Kuh 1998; Deecke et al. 2000; Campbell et al. 2002; Bahoura & Simard 2010; Charrier et al. 2010; Marcoux et al. 2011), Gunnison's prairie dogs, *Cynomys gunnisoni* (Placer & Slobodchikoff 2004), bats (Armitage & Ober 2010), grasshoppers (Chesmore & Ohya 2004), tungara frogs, *Engystomops pustulosus* (Phelps & Ryan 2000) and various bird species (Chesmore 2001; Terry & McGregor 2002; Aubin et al. 2004). For primates, Pozzi et al. (2010, 2012) showed their potential use in call type recognition and species recognition in lemurs, with an ANN approach achieving recognition accuracies of 94% for seven different call types within one species (*Eulemur macaco*) and 89% between different species of *Eulemur*.

In this study, we investigated whether a combination of ANN and MFCC could identify blue monkey, *Cercopithecus mitis stuhlmanni*, males individually by one call type, the 'pyow' alarm call (Papworth et al. 2008) and whether the same algorithm could also be applied to discriminate different call types in blue monkeys, and between 'pyows' and the calls of sympatric primate species. Besides showing the impact that software based on this method could have on research with primates, which feature strongly in behavioural research and conservation, this study is the first to attempt all three recognition tasks (individual, call type and species) with one set of parameters and the same classification tool. Previous research has mostly focused on single recognition tasks. We tested whether MFCC and neural networks can be the basis for a more generalized recognition software. This is an important step towards an equivalent of human speech and speaker recognition software in animal research.

METHODS

Study Site and Species

The study was carried out between May and August 2012 in one of the last remaining fragments of the Albertine Rift Forest Complex, the Budongo Forest Reserve in Masindi District, western Uganda (between 1°35′–1°55′N and 31°18′–31°42′E). The reserve comprises 793 km^2 of forest reserve, 428 km^2 of which are continuous forest cover (Fairgrieve & Muhumuza 2003), classified as moist semideciduous tropical forest at medium altitude (Plumptre & Reynolds 1994). In an initial 2-week period (18 May to 2 June 2012), the territories of all blue monkey groups in the study area were mapped using GPS and the grid system of the study area, to permit individualized data collection. In the 9 km^2 study area (Schel & Zuberbühler 2012), a total of 52 groups were identified, indicating a high group density of approximately 5.77 groups/km^2.

Like other forest guenons, blue monkeys live in one-male, multifemale groups with male dispersal. Groups are territorial, even though there is overlap between home ranges (Cords 2007). In blue monkeys, the 'pyow' call, a long-distance vocalization used exclusively by resident males, has been shown to differ between individuals (see Fig. 1 for two examples; Marler 1973; Butynski et al. 1992; Price et al. 2009). Acoustically, 'pyows' can be described as loud, explosive calls lasting on average about 110 ms, given in repetitive bouts (Marler 1973). The calls are given in response to a variety of events, mainly general disturbances and dangers on the ground, but also without apparent external stimulus (Butynski et al. 1992; Papworth et al. 2008; Murphy et al. 2013).

Data Collection

To train an ANN classifier, the identity of every individual included in the training set must be known in advance ('label validity': Clemins & Johnson 2003). Owing to the study species' social structure, 'pyow' calls recorded in a particular part of the forest could reliably be assigned to one specific male, the single male of the resident group, which is useful when training the ANN classifier. If group identity was in doubt, for example if the recording was made in an overlapping region, the group was followed for a while and neighbouring groups were located, to prevent misclassification.

Nineteen blue monkey groups were selected in order to obtain repeated recordings of male calls. 'Pyows' were obtained both through playback experiments and opportunistically. All vocal behaviour was recorded using a Marantz PMD-660 (D&M Holdings