Available online at www.sciencedirect.com

ScienceDirect

A new perspective on acoustic individual recognition in animals with limited call sharing or changing repertoires

ELIZABETH J. S. FOX

School of Animal Biology and School of Computer Science and Software Engineering,
University of Western Australia

The identification of individual animals based on acoustic parameters is a noninvasive method of recognizing individuals with considerable advantages over physical marking procedures which may be difficult to apply, time consuming, expensive or detrimental to the animal's welfare. To be an effective and practical method of individual recognition, an acoustic recognition technique must extract features that show greater variation between rather than within individuals and use a classifier that can successfully distinguish between the individuals and classify new recordings.

In addition, highly desirable features of an acoustic recognition technique include the following.

(1) The features exhibit little variation over time. This is necessary for studies requiring re-identification over time, with the required length that the features remain stable ranging from days to years, depending on the type of study.

(2) The classifier is able to determine when a feature set does not belong to any of the known individuals. This is important because animal populations are rarely closed, with new individuals arriving from immigration and births; hence a new recording may not belong to any of the known individuals and the classifier must be able to determine this.

(3) The features enable recognition regardless of the call type produced. This is important because recognition techniques that can compare only a single call type within and between individuals significantly limit the range of species and situations in which they can be used (N.B. The vocalizations of different species, and different types of vocalizations from the same species, often have specific descriptors: song, howl, call, etc. For simplicity, the term

call will be used in this paper to include all vocalization types, except when a particular species is being described in which case the correct term will be used).

Methods such as discriminant function analysis (DFA), using frequency and temporal measures, and spectrographic cross-correlation have demonstrated that individually distinctive calls are present in a wide range of species across many taxa and can be used to correctly identify individuals (Sparling & Williams 1978; Smith et al. 1982; McGregor et al. 2000; Osiejuk 2000). Individualistic calls most likely exist in all vocal animals as a result of genetic, developmental and environmental factors, although the level of individuality and whether it can be easily measured and classified will differ between species (Terry et al. 2005). Some studies have shown that vocal features can remain stable over days and even years (e.g. Lengagne 2001; Walcott et al. 2006), although there have been few extensive studies in this area. In addition, classification methods that are based on a similarity score, e.g. cross-correlation or adaptive kernel-based DFA, enable recognition of new individuals that have not been previously encountered (Terry et al. 2005). However, all of the current methods of acoustic recognition base the similarity of two vocalizations on a comparison of call-type-specific features (e.g. the frequency or length of a particular note or syllable). Hence comparisons both within and between individuals can occur only when the same call types are present, i.e. call-dependent recognition. Call-dependent recognition techniques therefore cannot be used or can be used only with difficulty under the following common conditions.

(1) Individuals temporarily change their calls. Temporary changes to a call involve short-term changes, usually in the frequency or temporal characteristics, of a particular call type and are a direct result of specific circumstances. Factors that have been shown to influence call characteristics include social context (Jones et al. 1993; Elowson & Snowdon 1994; Mitani & Brandt 1994), body condition

Correspondence: E. J. S. Fox, School of Animal Biology M092, University of Western Australia, 35 Stirling Hwy, Crawley, Western Australia 6009, Australia (email: foxe02@student.uwa.edu.au).

(Galeotti et al. 1997; Martin-Vivaldi et al. 1998; Poulin & Lefebvre 2003), time of year (Gilbert et al. 1994), emotional state (Bayart et al. 1990) and temperature (Friedl & Klump 2002). Temporary changes to calls probably occur in most animals. When identifying individuals from their calls, knowledge of the specific circumstances and how they affect the calls is required so that the affected variables can be excluded from analysis. For example, water temperature affects the temporal properties of European treefrog, *Hyla arborea*, calls (Friedl & Klump 2002) and hence temporal characteristics cannot be used to identify individuals over time. If this information is not known it may result in the variation present in the calls of an individual being greater between than within recordings, and this will result in incorrect identification.

(2) Individuals permanently change their calls. Permanent changes to a call usually involve the creation of new notes, syllables or entire calls, although they can also involve changes to the characteristics (e.g. frequency or temporal properties) of a particular call type. Permanent changes can be the result of a specific influencing factor or they can be a natural progression. An example of an influencing factor was found by Walcott et al. (2006) who showed that male loons, *Gavia immer*, have a yodel call that is stable from year to year but alters (in frequency and temporal properties) when the bird moves territory. A natural progression or continual change of call types is most commonly found in the oscine birds that are open-ended song learners or mimics. These birds incorporate new songs and calls into their repertoires throughout their lives. For example, noisy scrub-birds, *Atrichornis clamosus*, continually alter their song types over time, with significant changes in as little as 1 month and a complete repertoire change in 6 months (Berryman 2003). Other examples of species that change their repertoires over time include yellow-rumped caciques, *Cacicus cela* (Trainer 1989), boblinks, *Dolichonyx oryzivorus* (Avery & Oring 1977), pied flycatchers, *Ficedula hypoleuca* (Espmark & Lampe 1993) and superb lyrebirds, *Menura novaehollandiae* (Robinson & Curtis 1996). Permanent changes to call types are also found in young animals that must change from their immature begging calls to adult calls, often through a period of learning and experimentation (Kroodsma et al. 1982). Permanent changes to calls are likely to occur over longer time periods than temporary changes. The majority of studies examining acoustic identification have used calls recorded over a short time period, usually within a single breeding season (Otter 1996; Hill & Lill 1998; McCowan & Hooper 2002; Rogers & Paton 2005). Markedly fewer studies have been carried out on the stability of vocalizations between years (Lengagne 2001; Gilbert et al. 2002; Puglisi & Adamo 2004).

(3) Individuals in a species have limited call sharing. Animal populations can vary in the number of calls that are shared between individuals, from complete sharing of all call types to active avoidance of call sharing (Catchpole & Slater 1995). The amount of call sharing also depends on the distance over which individuals are studied. Neighbouring birds may have extensive call sharing, but there is a decrease in sharing with an increase in spatial separation

in many species (e.g. Farabaugh et al. 1988; Rogers 2002). Having limited call sharing between individuals creates two problems. First, a separate classifier must be created for each call type that is shared between individuals. This can lead to a large number of classifiers being required if each call type is shared between only a small number of individuals. For example, of 38 song types sung by six male rufous bristlebirds, *Dasyornis broadbenti*, the most common song types were shared between only four of the six individuals (Rogers & Paton 2005). To distinguish between all six birds it was therefore necessary to carry out classifications on several song types, with each classification able to distinguish between only two and four birds. This makes the method very time consuming because a classifier has to be created for each call type. In addition, each recording must be separated into its respective call types before analysis and classification can occur, which can be a particularly arduous task for species with large repertoires. Second, it is necessary to know the complete set of calls from each individual. Without knowledge of the complete repertoire from each individual, a novel call may be incorrectly attributed to a new bird in the population. Limited call sharing is found in many oscine species, e.g. Kentucky warblers, *Oporornis formosus* (Tsipoura & Morton 1988), rufous bristlebirds (Rogers 2004), dark-eyed juncos, *Junco hyemalis* (Williams & MacRoberts 1978) and song sparrows, *Melospiza melodia* (Borror 1965).

(4) Individuals have extensive repertoires and/or use repeat-mode calling. About 70% of songbirds produce multiple song types (Beecher & Brenowitz 2005). These repertoires range in size from less than five songs, e.g. great tits, *Parus major* to over 1000, e.g. brown thrashers, *Toxostoma rufum* (Beecher & Brenowitz 2005). When an individual has a large repertoire, long recordings may be needed before the particular song required to determine identity is obtained. The recording length required can be even longer if the species is a repeat-mode caller (Wiley et al. 1994) in which only a single song type is repeated within a bout of singing (e.g. rufous bristlebirds; Rogers & Paton 2005). It may therefore be hours or days before the required song type is produced and recorded, making acoustic recognition based on the comparison of a particular call type a long, arduous and manually intensive exercise.

It is clear that with only call-dependent recognition, acoustic individual recognition is limited to species with extensive call sharing and no change in an individual's repertoire over time. The most common animal taxa that do not obey these requirements are the passerine, and particularly the oscine, bird species. The inability of current methods to work successfully with these species is demonstrated by the fact that, although there are roughly twice as many passerines as nonpasserines (Pimm et al. 2006), a recent literature search found that of 53 published studies on acoustic individual recognition in birds only 30% were carried out on passerine species. Other animals to which call-dependent recognition is applicable only in a limited way include mammals with complex calling systems such as cetaceans and primates.

Current methods of acoustic recognition are call dependent because they require the comparison of features that are specific to a particular call type. To carry out acoustic recognition regardless of call type, features that are specific to the individual's voice and remain stable regardless of the particular call produced must be found. It is well known that humans can easily recognize other people from their voices and this has led to the development of speaker recognition technology. Initial approaches at identifying people from their voice characteristics used long-term averaged features (Markel et al. 1977). Similar techniques were tested on great tits by Weary et al. (1990) who used long-term-averaged temporal and frequency features across different song types, resulting in a recognition accuracy of 69.9–80.4%. Long-term averaging of features is an extreme condensation of the characteristics of the voice and discards a lot of individual information (Reynolds 1995). Hence speaker recognition technology currently uses short-term features that are extracted from 10- to 30-ms segments of the signal. These features are based on the characteristics of the vocal tract shape and are therefore specific to the individual, not to the particular words spoken. These short-term features have been used with great success, resulting in speaker recognition accuracies of typically 80–100% (e.g. Farrell et al. 1994; Matsui & Furui 1994; Reynolds & Rose 1995; Murthy et al. 1999). In recent years researchers have begun to apply these same methods to the problem of animal individual recognition. In the African elephant, *Loxodonta africana*, 82.5% individual recognition accuracy was achieved (Clemins et al. 2005), while in the Norwegian ortolan bunting, *Emberiza hortulana*, Trawicki et al. (2005) identified 80–95% of individuals correctly. These were both call-dependent recognition tasks in which only a single call type was compared. One of the major advantages that speaker recognition techniques can bring to individual recognition in animals is the ability for recognition regardless of call type, i.e. call-independent recognition.

Speaker Recognition Methods

I will briefly discuss the methods of feature extraction and classification commonly used in speaker recognition and then present the results of some preliminary tests using these methods to demonstrate that they are a feasible method of call-independent individual recognition in a passerine species. My major aim is to demonstrate a new approach to individual recognition using acoustic cues that overcomes most of the limitations of current approaches. I present one example to show that the methods have real potential. Its application more broadly can be evaluated only by rigorous application in a variety of animals using acoustic signals.

Speaker recognition is a topic within the field of speech processing and refers to the ability to identify an individual based on aspects of their voice (Farrell 2000). When only a single set of text (i.e. words or sentences) is used for both training and testing a classifier recognition is termed text dependent. When the text varies between training and testing recognition is termed text independent (Furui 1997). The ability to carry out text-independent recognition lies in the selection of acoustic features that remain relatively stable regardless of the sounds produced. In humans, voiced sound is produced by the vibration of the vocal cords, which results in a quasi-periodic flow of air called the source sound (Masaki 2000). This source sound is characterized by its fundamental frequency and harmonic overtones, which are determined by the subglottal pressure, and the tension of the vocal cords. The source sound passes through the vocal tract, consisting of the nasal and oral cavities in association with the lips, tongue, jaw and teeth (Furui 2001), which alters the frequency content through a modulation of the amplitude of the harmonics. The modulation is a result of the resonances of the vocal tract, which are a consequence of the size and shape of the vocal tract. The resulting spectral shape, called formants (Fig. 1), can be measured from a signal and from this the individual's vocal tract shape can be estimated.
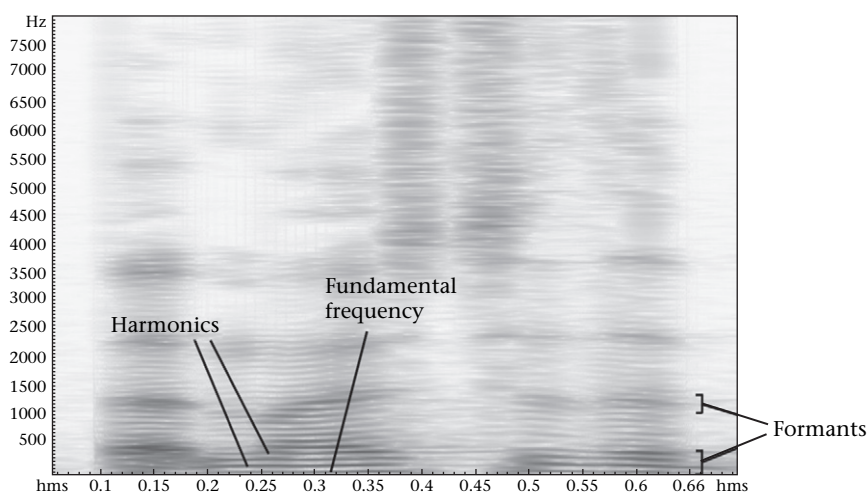

Figure 1. Spectrogram of a speech segment.

This idea of sound production is approximated by the source-filter model of speech production (Fig. 2),

$$y(n) = s(n) * h(n),$$

where $y(n)$ is the speech signal in the time domain and $s(n)$ is the source sound that is convolved with $h(n)$, the vocal tract filter. Although this model was developed for human speech, it can be applied to any sound that is produced at a source and then modified by a filter. For example, mammalian and avian vocal production (Lieberman 1969; Nowicki & Marler 1988) and musical instruments (Eronen 2001) can be modelled by the source-filter model.

For human speech, features of the sound that result from the vocal tract resonances contain the most individually specific information. It is therefore necessary to separate the vocal tract and source sound information. These features are convolved with each other in the spectral domain and cannot be separated, but, through the use of homomorphic analysis, the signal can be converted to the cepstral domain where the source and vocal tract features are no longer convolved and can be easily separated from each other (Furui 2001; Quatieri 2002),

$$Y(n) = S(n) + H(n),$$

where $Y(n)$, $S(n)$ and $H(n)$ are the signal, source sound and vocal tract filter in the cepstral domain. The term cepstral is derived from the word spectral because the cepstral domain is the inverse Fourier transform of the logarithmic amplitude spectrum of a signal (Furui 2001).

In the cepstral domain the lower-order coefficients represent the spectral envelope (the vocal tract information) while the source information is represented in the higher-order coefficients. Therefore, typically only the first 12−15 cepstral coefficients are used (Gish & Schmidt 1994).

The most common features used for human speaker identification are the mel-frequency cepstral coefficients (Campbell 1997; Quatieri 2002) developed by Davis & Mermelstein (1980). These cepstral coefficients are calculated using a filterbank based on the mel-scale of frequencies. The mel-scale approximates the human perception of frequency, which follows a logarithmic rather than linear scale above 1 kHz (Mammone et al. 1996). The mel-frequency cepstral coefficients (MFCCs) are popular because they tend to be uncorrelated, are computationally efficient, incorporate human perceptual information and have some resilience to noise (Quatieri 2002; Clemins 2005), all of which result in higher recognition accuracies. Recently there has been interest in using perceptual lin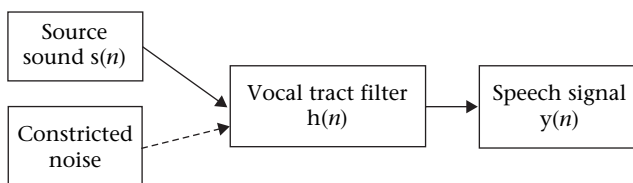ear prediction (PLP) coefficients, particularly for nonhuman species, because PLP analysis can incorporate information about the auditory ability of the species under study (Clemins & Johnson 2006). The PLP model was developed by Hermansky (1990) and stresses perceptual accuracy over computational efficiency. The generalized PLP developed by Clemins & Johnson (2006) enables human perceptual information to be replaced with species-specific information which may lead to improved recognition accuracy in nonhuman species.

Once individually specific features have been extracted, a classifier that can be trained to distinguish between the feature sets and then can test a new feature set by comparing it with the stored reference templates for each individual to make a decision about identity is required (Farrell 2000; Furui 2001; Ramachandran et al. 2002). Some common classifiers used for speaker recognition include dynamic time warping, hidden Markov models, Gaussian mixture models and artificial neural networks (Furui 1997; Ramachandran et al. 2002). The type of classifier used depends on the required task. Some classifiers, such as dynamic time warping and hidden Markov models, include temporal information and therefore are best suited to text-dependent recognition, while others, such as Gaussian mixture models and artificial neural networks, have shown good results for text-independent tasks (Ramachandran et al. 2002).

Below I demonstrate the potential for call-independent individual recognition in willie wagtails, *Rhipidura leucophrys*, using mel-frequency cepstral coefficients and an artificial neural network.

Experimental Methods

The songs of 10 willie wagtails were recorded from locations around Perth, Western Australia using a Sony ECM672 directional microphone with a Marantz PMD670 solid-state recorder at a sampling frequency of 48 kHz. Birds were recorded at night (2000 to 0400 hours) during spring, at which time wagtails frequently sit in a single location and sing for long periods. All recordings were initially analysed using Cool Edit Pro (Syntrillium Software Corp). The silent (nonsong) parts of the recordings were removed through the use of an amplitude filter and each recording was high-pass filtered at 700 Hz to remove low-frequency background noise. Each recording was then split into its respective song types through a visual inspection of the spectrograms. One song type was used for training the classifier, and a different song type was used to test the classifier (Fig. 3). Training was carried out using 10 s of recording, plus 10 s were used as a validation set to enable early stopping which prevents the network from overtraining and losing the ability to generalize. Ten, 1-s tests were carried out for each individual on the trained network using the second song type. For both the training and the testing data, the 12th-order MFCCs were extracted from 30-ms frames and fed to the classifier. The classifier used was an artificial neural network, a multilayer perceptron (MLP), which was designed and implemented using the neural network toolbox in Matlab (version 6.5.1; The MathWorks, Inc.,). The network had one hidden layer with 16 neurons.
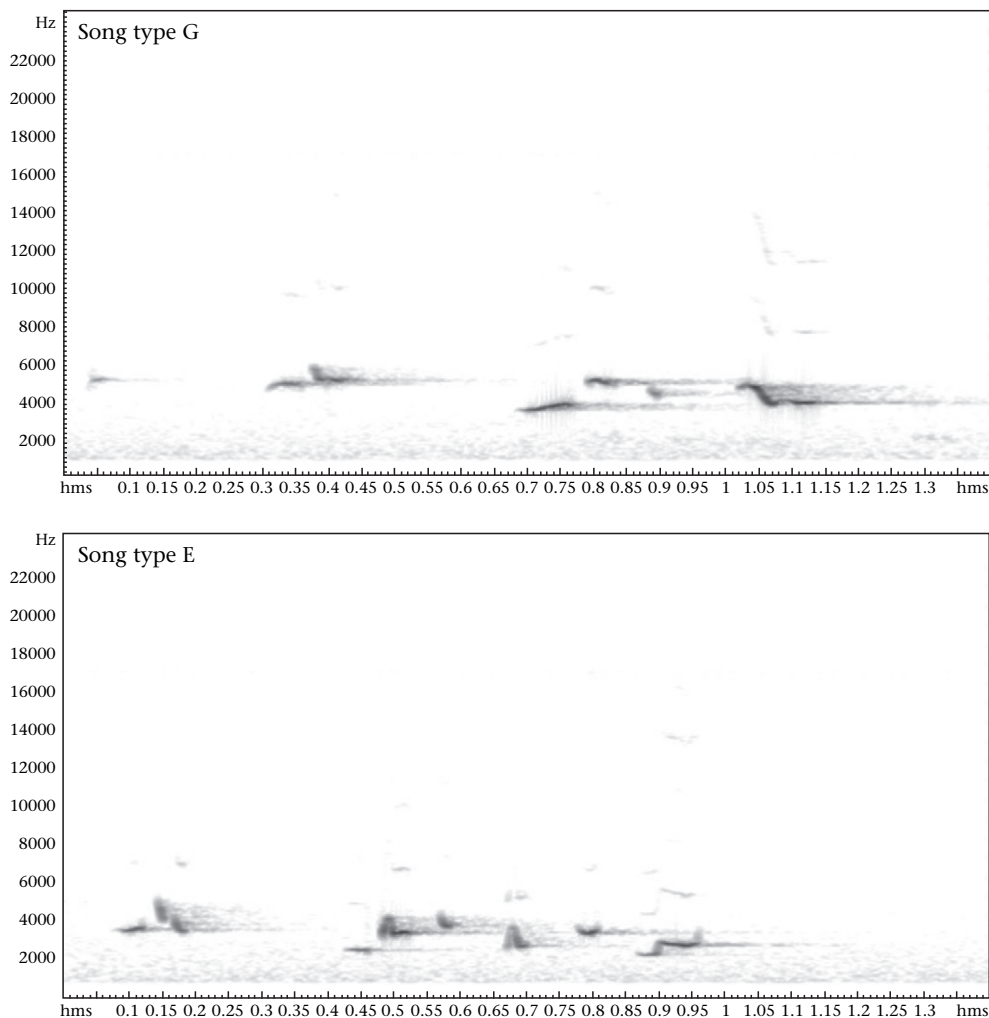

Figure 2. Source-filter model of speech production.

Figure 3. Example of the different song types used for training and testing for a single wagtail.

Results and Discussion

Call-independent recognition in willie wagtails using MFCCs and an MLP resulted in a recognition accuracy of 89%. The confusion matrix of the results is shown in Table 1,

Table 1. Confusion matrix of testing and training with different song types (e.g. 2C = bird 2, song type C)

		Training									
		2C	3S	8E	9G	10E	17G	20A	21E	27E	30E
	2A	10	0	0	0	0	0	0	0	0	0
	3R	0	6	0	0	0	3	1	0	0	0
	8G	0	0	10	0	0	0	0	0	0	0
	9E	0	0	0	9	1	0	0	0	0	0
	10F	0	0	0	0	10	0	0	0	0	0
Testing	17A	0	0	2	1	0	7	0	0	0	0
	20C	0	0	0	0	0	0	10	0	0	0
	21A	0	0	0	0	0	0	0	10	0	0
	27G	0	0	0	0	0	0	0	0	7	3
	30F	0	0	0	0	0	0	0	0	0	10

with the identity and song type of each bird trained running horizontally and the identity and song type of each bird tested running vertically. The results of the 10 tests carried out for each bird are placed under the bird and song type to which the MLP classified them. Call-independent recognition is typically more difficult than call-dependent recognition, so the high result achieved in this call-independent task, which is comparable to the result for call-dependent recognition in the Norwegian ortolan bunting (Trawicki et al. 2005), is particularly encouraging.

The fact that the cepstral coefficients are extracting features of the voice, rather than features specific to the song type, was demonstrated in the tests in which a single song type was used for both training and testing in different individuals (for example song type A was used for training in bird 20 and used for testing in bird 2). In 69 of the 70 tests in which the same song type was used for training and testing in different individuals, the song type was successfully classified to the correct individual, rather than to the same song type.

This experiment used methods of feature extraction and classification taken directly from human speaker

recognition tasks. It is likely that the results can be improved by modifying the methods to better suit bird song or by using methods specifically designed to incorporate species-specific information (for example the generalized PLP model of Clemins & Johnson 2006). In addition, because the same methods give good results for both human speech and bird song, it is likely that these methods can be used across a wide range of species.

All identification techniques contain limitations and potential biases which must be taken into account before choosing the correct method for each species or type of study. As with any method of acoustic individual recognition, the study population is limited to those individuals that produce vocalizations, which may be affected by factors such as sex, age or breeding status (Terry et al. 2005). Another potential limitation is that the extraction of features through speaker recognition methods, such as cepstral analysis, is based upon the source-filter model of sound production. Not all animal sounds are produced in this way, for example the clicks and noises produced by some cetaceans (Cranford et al. 1996) or the sounds produced by insects (Alexander 1957). However, these sounds are likely to contain individual characteristics and speaker recognition methods may still provide useful information. For example, cepstral analysis improved species identification in crickets, katydids and cicadas (Ganchev et al. 2007). Individual recognition using speaker recognition techniques has currently been studied in only a few species, although the successful application of the same methods to species exhibiting a range of vocalization frequencies and abilities, including elephants (Clemins et al. 2005), pigs (Schon et al. 2001), and a passerine (Trawicki et al. 2005), implies that the methods are widely applicable. Studies on species with differing sound production methods and types of vocalizations, e.g. frogs, cetaceans or insects, will be necessary before the full extent of the application of speaker recognition methods can be determined.

Another potential problem with using speaker recognition techniques on field recordings of animals is that noise, and in particular the mismatched conditions that occur when a recording used for testing a classifier has noise different from that with which the classifier was trained, is known to be a major challenge in human speaker recognition applications (Juang 1991). Noise can arise from a variety of sources such as ambient noise, reverberations, channel interference or microphone distortions. Although excellent recognition performance can be achieved when the recording conditions are matched between training and testing, a dramatic drop in accuracy can occur under mismatched conditions. For example a 10-dB addition of Gaussian noise was seen to decrease accuracy by up to 80% when identifying human voices (Gong 1995). There are many noise removal methods that can increase this accuracy to less than 20% below that obtained for matched recordings (Gong 1995). It is likely that background noise and signal degradation will be a significant problem for animal acoustic recognition due to the variable nature of weather conditions, other background noise and distance from the subject that are inherent in obtaining field recordings. The recordings used in this experiment had little background noise because they were obtained at night-time and with the microphone usually within 5 m of the bird. Because birds are often recorded during the dawn chorus, there will typically be much greater levels of background noise and it may be harder to approach the birds closely. Effort may need to be spent researching the impact of noise and other distortions before the techniques outlined above become generally applicable to field situations.

Conclusion

Acoustic individual recognition has the potential to be an extremely useful tool for studying individual behaviours and in ecological contexts requiring individual recognition. It has the advantage over physical marking techniques of being noninvasive, inexpensive and relatively fast and simple to apply. Developing a method of call-independent recognition will, for the first time, provide a method of individual recognition that can be applied to all species regardless of the complexity of calls, amount of call sharing or individual variation in calls over time. In addition, speaker recognition techniques solve several of the other problems associated with the current methods of acoustic individual recognition, which has resulted in their rare use as methods of individual recognition, including.

(1) The classifiers enable new calls to be classified as unknown individuals.

(2) The methods are not species specific, thereby preventing the need for extensive pilot studies.

(3) Call-independent recognition eliminates the need to separate recordings into their respective song types, thereby saving considerable amounts of time and effort.

(4) Feature extraction and classification are both carried out automatically, again resulting in a saving of time and effort.

Conveniently, human speaker recognition techniques appear to be just as applicable to animal vocalizations as to human speech and hopefully research in this area will result in substantial improvements in the ease and way in which animals are studied.

References

Alexander, R. D. 1957. Sound production and associated behavior in insects. *Ohio Journal of Science*, **57**, 101–113.

Avery, M. & Oring, L. W. 1977. Song dialects in the boblink (*Dolichonyx oryzivorus*). *Condor*, **79**, 113–118.

Bayart, F., Hayashi, K. T., Faull, K. F., Barchas, J. D. & Levine, S. 1990. Influence of maternal proximity on behavioral and physiological responses to separation in infant rhesus monkeys. *Behavioral Neuroscience*, **104**, 98–107.

Beecher, M. D. & Brenowitz, E. A. 2005. Functional aspects of song learning in songbirds. *Trends in Ecology & Evolution*, **20**, 143–149.

Berryman, A. N. 2003. Can consistent individuality of voice be used to census the vulnerable noisy scrub-bird *Atrichornis clamosus*? Honours thesis, Murdoch University, Western Australia.

Borror, D. J. 1965. Song variation in Maine song sparrows. *Wilson Bulletin*, **77**, 5–37.

Campbell, J. P. 1997. Speaker recognition: a tutorial. *Proceedings of the IEEE*, **85**, 1437–1462.

Catchpole, C. K. & Slater, P. J. B. 1995. *Bird Song: Biological Themes and Variations*. Cambridge: Cambridge University Press.

Clemins, P. J. 2005. Automatic classification of animal vocalizations. Ph.D. thesis, Marquette University.

Clemins, P. J. & Johnson, M. T. 2006. Generalized perceptual linear prediction features for animal vocalization analysis. *Journal of the Acoustical Society of America*, **120**, 527–534.

Clemins, P. J., Johnson, M. T., Leong, K. M. & Savage, A. 2005. Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations. *Journal of the Acoustical Society of America*, **117**, 1–8.

Cranford, T. W., Amundin, M. & Norris, K. S. 1996. Functional morphology and homology in the odontocete nasal complex: implications for sound generation. *Journal of Morphology*, **228**, 223–285.

Davis, S. B. & Mermelstein, P. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **28**, 357–366.

Elowson, A. M. & Snowdon, C. T. 1994. Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment. *Animal Behaviour*, **47**, 1267–1277.

Eronen, A. 2001. Comparison of features for musical instrument recognition. In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 19–22.

Espmark, Y. O. & Lampe, H. M. 1993. Variations in the song of the pied flycatcher within and between breeding seasons. *Bioacoustics*, **5**, 33–65.

Farabaugh, S. M., Brown, E. D. & Veltman, C. J. 1988. Song sharing in a group-living songbird the Australian magpie. Part II. Vocal sharing between territorial neighbors within and between geographic regions and between sexes. *Behaviour*, **104**, 105–125.

Farrell, K. R. 2000. Networks for speaker recognition. In: *Handbook of Neural Networks for Speech Processing* (Ed. by S. Katagiri), pp. 357–391. Norwood: Artech House.

Farrell, K. R., Mammone, R. J. & Assaleh, K. T. 1994. Speaker recognition using neural networks and conventional classifiers. *IEEE Transactions on Speech and Audio Processing*, **2**, 194–205.

Friedl, T. W. P. & Klump, G. M. 2002. The vocal behaviour of male European treefrogs (*Hyla arborea*): implications for inter- and intrasexual selection. *Behaviour*, **139**, 113–136.

Furui, S. 1997. Recent advances in speaker recognition. *Pattern Recognition Letters*, **18**, 859–872.

Furui, S. 2001. *Digital Speech Processing, Synthesis, and Recognition*. New York: Marcel Dekker.

Galeotti, P., Saino, N., Sacchi, R. & Moller, A. P. 1997. Song correlates with social context, testosterone and body condition in male barn swallows. *Animal Behaviour*, **53**, 687–700.

Ganchev, T., Potamitis, I. & Fakotakis, N. 2007. Acoustic monitoring of singing insects. In: *ICASSP*.

Gilbert, G., McGregor, P. K. & Tyler, G. 1994. Vocal individuality as a census tool: practical considerations illustrated by a study of two rare species. *Journal of Field Ornithology*, **65**, 335–348.

Gilbert, G., Tyler, G. A. & Smith, K. W. 2002. Local annual survival of booming male Great Bittern *Botaurus stellaris* in Britain, in the period 1990–1999. *Ibis*, **144**, 51–61.

Gish, H. & Schmidt, M. 1994. Text-independent speaker identification. *IEEE Signal Processing Magazine*, **11**, 18–31.

Gong, Y. 1995. Speech recognition in noisy environments: a survey. *Speech Communication*, **16**, 261–291.

Hermansky, H. 1990. Perceptual linear predictive (PLP) analysis of speech. *Journal of the Acoustical Society of America*, **87**, 1738–1752.

Hill, F. A. R. & Lill, A. 1998. Vocalisations of the Christmas Island hawk-owl *Ninox natalis*: individual variation in advertisement calls. *Emu*, **98**, 221–226.

Jones, B. S., Harris, D. H. R. & Catchpole, C. K. 1993. The stability of the vocal signature in phee calls of the common marmoset, *Callithrix jacchus*. *American Journal of Primatology*, **31**, 67–75.

Juang, B. H. 1991. Speech recognition in adverse environments. *Computer Speech and Language*, **5**, 275–294.

Kroodsma, D. E., Miller, E. H. & Ouellet, H. 1982. *Acoustic Communication in Birds*. New York: Academic Press.

Lengagne, T. 2001. Temporal stability in the individual features in the calls of eagle owls (*Bubo bubo*). *Behaviour*, **138**, 1407–1419.

Lieberman, P. 1969. On the acoustic analysis of primate vocalizations. *Behavioral Research, Methods, and Instrumentation*, **1**, 169–174.

McCowan, B. & Hooper, S. L. 2002. Individual acoustic variation in Belding's ground squirrel alarm chirps in the High Sierra Nevada. *Journal of the Acoustical Society of America*, **111**, 1157–1160.

McGregor, P. K., Peake, T. M. & Gilbert, G. 2000. Communication behaviour and conservation. In: *Behaviour and Conservation* (Ed. by L. M. Gosling & W. J. Sutherland), pp. 261–280. Cambridge: Cambridge University Press.

Mammone, R. J., Zhang, X. Y. & Ramachandran, R. P. 1996. Robust speaker recognition: a feature-based approach. *IEEE Signal Processing Magazine*, **13**, 58–71.

Markel, J. D., Oshika, B. T. & Gray, A. H. 1977. Long-term feature averaging for speaker recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **25**, 330–337.

Martin-Vivaldi, M., Palomino, J. J. & Soler, M. 1998. Song structure in the hoopoe (*Upupa epops*): strophe length reflects male condition. *Journal of Ornithology*, **139**, 287–296.

Masaki, S. 2000. The speech signal and its production model. In: *Handbook of Neural Networks for Speech Processing* (Ed. by S. Katagiri), pp. 19–62. Norwood: Artech House.

Matsui, T. & Furui, S. 1994. Comparison of text-independent speaker recognition methods using VQ-distortion and discrete/continuous HMMs. *IEEE Transactions on Speech and Audio Processing*, **2**, 456–459.

Mitani, J. C. & Brandt, K. 1994. Social factors influence the acoustic variability in the long-distance calls of male chimpanzees. *Ethology*, **96**, 233–252.

Murthy, H. A., Beaufays, F., Heck, L. P. & Weintraub, M. 1999. Robust text-independent speaker identification over telephone channels. *IEEE Transactions on Speech and Audio Processing*, **7**, 554–568.

Nowicki, S. & Marler, P. 1988. How do birds sing? *Music Perception*, **5**, 391–426.

Osiejuk, T. S. 2000. Recognition of individuals by song, using cross-correlation of sonograms of Ortolan buntings *Emberiza hortulana*. *Biological Bulletin of Poznan*, **37**, 95–106.

Otter, K. 1996. Individual variation in the advertising call of male northern saw-whet owls. *Journal of Field Ornithology*, **67**, 398–405.

Pimm, S., Raven, P., Peterson, A., Sekercioglu, C. H. & Ehrlich, P. R. 2006. Human impacts on the rates of recent, present, and

future bird extinctions. *Proceedings of the National Academy of Sciences, U.S.A.,* **103**, 10941–10946.

Poulin, B. & Lefebvre, G. 2003. Variation in booming among great bitterns *Botaurus stellaris* in the Camargue, France. *Ardea,* **91**, 177–181.

Puglisi, L. & Adamo, C. 2004. Discrimination of individual voices in male great bitterns (*Botaurus stellaris*) in Italy. *Auk,* **121**, 541–547.

Quatieri, T. F. 2002. *Discrete-time Speech Signal Processing: Principles and Practice.* Upper Saddle River, New Jersey: Prentice Hall.

Ramachandran, R. P., Farrell, K. R., Ramachandran, R. & Mammone, R. J. 2002. Speaker recognition: general classifier approaches and data fusion methods. *Pattern Recognition,* **35**, 2801–2821.

Reynolds, D. A. 1995. Large population speaker identification using clean and telephone speech. *IEEE Signal Processing Letters,* **2**, 46–48.

Reynolds, D. A. & Rose, R. C. 1995. Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing,* **3**, 72–83.

Robinson, F. N. & Curtis, H. S. 1996. The vocal displays of the lyrebirds (Menuridae). *Emu,* **96**, 258–275.

Rogers, D. 2002. Intraspecific variation in the acoustic signals of birds and frogs: implications for the acoustic identification of individuals. Ph.D. thesis, University of Adelaide.

Rogers, D. 2004. Repertoire size, song sharing and type matching in the rufous bristlebird (*Dasyornis broadbenti*). *Emu,* **104**, 7–13.

Rogers, D. J. & Paton, D. C. 2005. Acoustic identification of individual rufous bristlebirds, a threatened species with complex song repertoires. *Emu,* **105**, 203–210.

Schon, P.-C., Puppe, B. & Manteuffel, G. 2001. Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*). *Journal of the Acoustical Society of America,* **110**, 1425–1431.

Smith, H. J., Newman, J. D., Hoffman, H. J. & Fetterly, K. 1982. Statistical discrimination among vocalizations of individual squirrel monkeys (*Saimiri sciureus*). *Folia Primatologia,* **37**, 267–279.

Sparling, D. W. & Williams, J. D. 1978. Multivariate analysis of avian vocalizations. *Journal of Theoretical Biology,* **74**, 83–107.

Terry, A. M. R., Peake, T. M. & McGregor, P. K. 2005. The role of vocal individuality in conservation. *Frontiers in Zoology,* **2**.

Trainer, J. M. 1989. Cultural evolution in song dialects of yellow-rumped caciques in Panama. *Ethology,* **80**, 190–204.

Trawicki, M. B., Johnson, M. T. & Osiejuk, T. S. 2005. Automatic song-type classification and speaker identification of Norwegian ortolan bunting. In: *IEEE International Conference on Machine Learning in Signal Processing (MLSP)*.

Tsipoura, N. & Morton, E. S. 1988. Song-type distribution in a population of Kentucky warblers. *Wilson Bulletin,* **100**, 9–16.

Walcott, C., Mager, J. N. & Walter, P. 2006. Changing territories, changing tunes: male loons, *Gavia immer*, change their vocalizations when they change territories. *Animal Behaviour,* **71**, 673–683.

Weary, D. M., Norris, K. J. & Falls, J. B. 1990. Song features birds use to identify individuals. *Auk,* **107**, 623–625.

Wiley, R. H., Godard, R. & Thompson, A. D. 1994. Use of two singing modes by hooded warblers as adaptations for signalling. *Behaviour,* **129**, 243–278.

Williams, L. & MacRoberts, M. H. 1978. Song variation in dark-eyed juncos in Nova Scotia. *Condor,* **80**, 237–240.