

Automatic identification of individual killer whales

Judith C. Brown^{a)}

*Department of Physics, Wellesley College, Wellesley, Massachusetts 02481 and Media Laboratory,
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
brown@media.mit.edu*

Paris Smaragdis

*Adobe Systems, Cambridge, Massachusetts 02139
paris@adobe.com*

Anna Nousek-McGregor

*Duke Marine Laboratory, Beaufort, North Carolina 28516
aem41@duke.edu*

Abstract: Following the successful use of HMM and GMM models for classification of a set of 75 calls of northern resident killer whales into call types [Brown, J. C., and Smaragdis, P., *J. Acoust. Soc. Am.* **125**, 221–224 (2009)], the use of these same methods has been explored for the identification of vocalizations from the same call type N2 of four individual killer whales. With an average of 20 vocalizations from each of the individuals the pairwise comparisons have an extremely high success rate of 80 to 100% and the identifications within the entire group yield around 78%.

© 2010 Acoustical Society of America

PACS numbers: 43.80.Ka, 43.80.Ev, 43.60.Uv [CM]

Date Received: May 6, 2010 **Date Accepted:** June 3, 2010

1. Introduction

The automatic identification of individual animals from the sounds they produce has been discussed recently by [Adi *et al.* \(2010\)](#) and applied to the Norwegian ortalan bunting for purposes of acoustic censusing. Previous work on marine mammal sounds was reported by [Nousek \(2004\)](#) and [Nousek *et al.* \(2006\)](#) doing pairwise comparisons of killer whale sounds of the same call type. As features they used the frequency contours of the calls as input to a neural network ([Deecke *et al.*, 1999](#)). Earlier results on marine mammals, also using frequency contours, were reported by [Buck and Tyack \(1993\)](#) in classifying 15 bottlenose dolphin *Tursiops truncatus* signature whistles into five groups using Dynamic Time Warping.

Our calculations report the first use of statistical methods to identify individual marine mammals from the time frequency decomposition of their sounds. These sounds were previously classified as belonging to the same call type so this calculation is comparable to the identification of humans based on the utterance of the same word or short phrase. They are from the same set of killer whale sounds used in the calculation by [Nousek \(2004\)](#) and [Nousek *et al.* \(2006\)](#) making possible a direct comparison of methodologies.

2. Background

2.1 Gaussian mixture models (GMM) and hidden Markov models (HMM)

The GMM method is used to perform classification or identification by evaluating the average model likelihood of spectra over the entire input utterance. Consequently the temporal structure of the sound is not taken into account. The HMM calculation detects both spectral and temporal

^{a)} Author to whom correspondence should be addressed.

sequences, that is the progression of spectral changes in the sound. Thus the GMM should pick up differences in the spectral qualities of individuals, and the HMM should be sensitive to temporal changes as well, for example, small frequency variations. Both methods have the advantage of being applicable directly to the time-frequency decomposition without requiring the tedious pre-processing step of obtaining the frequency contours. These methods are discussed more fully in recent work on classification of killer whale vocalizations, which includes a summary of their use in bioacoustics. (Brown and Smaragdis, 2008, 2009)

The Gaussian mixture model (GMM) is a commonly used estimate of the probability density function used in statistical classification and identification systems. Gaussian mixture models (GMM) have found widespread use in speech research, primarily for speaker recognition (Reynolds and Rose, 1995 and references therein) and have been used in other fields, for example for musical instrument identification (Brown, 1999 and Brown *et al.*, 2001), and, as mentioned, by Adi *et al.* (2010) for the Norwegian ortolan bunting. They have proven to be well suited for many and varied identification problems.

The Hidden Markov model treats the data as a sequence of *states*. The states can be considered as separate GMM's with their temporal evolution governed by a *transition matrix*. This matrix is learned from the training data and defines the probabilities of moving from one state to the next. In sum, the HMM creates a sequence of GMM models to explain the input data. It takes account of the temporal structure of the sound and uses it as additional information for identification.

The choice of cepstra as features has been particularly successful in characterizing the vocal tract resonances which identify individual human speakers, speech, or vowels. The cepstrum is the Fourier transform of the log magnitude spectrum (Oppenheim and Schaffer, 1975); it involves two transforms which makes it computationally more intensive than FFT based calculations. See Rabiner and Schaffer (1978) and Rabiner and Juang (1993) for a discussion of the use of cepstra for speech applications.

3. Calculations and results

Sounds from call type N2 were chosen for this calculation as they offered the advantage of 4 individuals with over 10 calls of each in the database. The features chosen for all of the calculations were cepstral coefficients and their temporal derivatives. These were calculated with the program melcepst available with the MATLAB toolbox VOICEBOX.¹ The number of cepstral coefficients was varied from 12 to 30 with best results for 24 and over so results are reported for 24 cepstra. The sample rate was 48000 samples/s with each sound divided into overlapping 10 ms segments for the calculations of both GMM and HMM models. The GMM/HMM computations were carried out with custom software written in MATLAB for this task. The training set for all classifications consisted of all the sounds except the one being classified, called the "leave one out" method.

3.1 Sounds

Killer whale sounds were recorded in 1998 and 1999 in Johnstone Strait, British Columbia, using methodology developed by Miller and Tyack (1998). Sounds of four individuals from the A clan previously classified as belonging to call type N2 were chosen for this calculation. See examples of spectrograms in Fig. 1. The individuals were A32 (16 sounds), A46 (14 sounds), A12 (11 sounds), and A8 (39 sounds). See Nousek (2004), Ford (1987), Miller and Bain (2000), and Brown (2008) for notation, examples and a mathematical discussion of spectra. Individuals A32 and A46 were members of the same matriline (A36) with A12 from the same pod (A1) but a different matriline (A12). Individual A8 was from a different pod (A5) and matriline (A8). Thus one might expect A32 and A46 to have the most similar sounds and those of A8 to be the most distinguishable from the other three.

3.2 GMM results

Results for the GMM calculations on the vocalizations of all four individuals together ("ALL") and for the six pairwise comparisons are given in Fig. 2 with the number of Gaussians in the

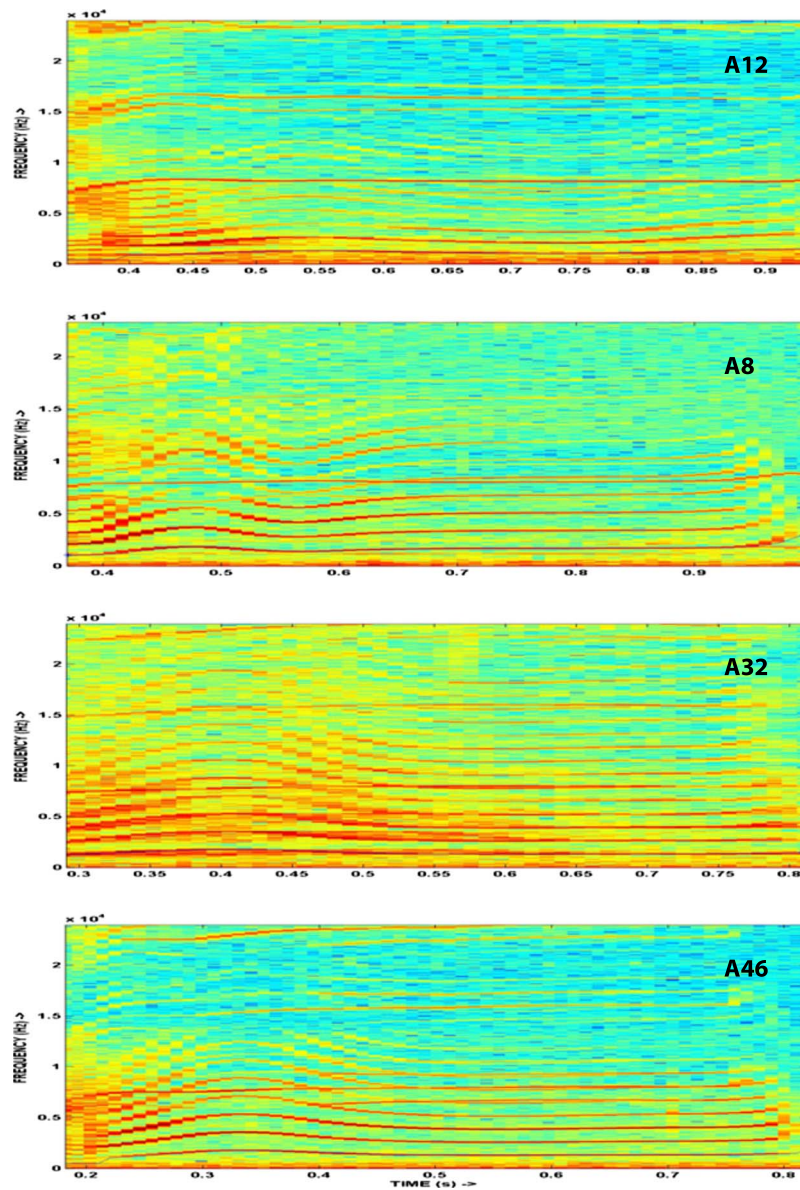


Fig. 1. (Color online) Examples of the spectrograms of N2 calls by killer whales A12, A8, A32, and A46, included in this study.

probability distributions varying from 1 to 8. Results are best for 6 or more Gaussians and range from 85 to 100% correct for the pairwise comparisons. The identification from all four individuals was over 75% which is excellent, noting that random selection would give 25%.

3.3 HMM results

For the HMM classification, a left-to-right model was used; and there were three variable parameters rather than two. The number of Gaussians in the probability function was varied from 1 to 3 with the results slightly better for two. The number of states was varied from 5 to 11 (Fig. 3), and with one exception there is less than a 5% variation depending on number of classes. This indicates a highly robust calculation.

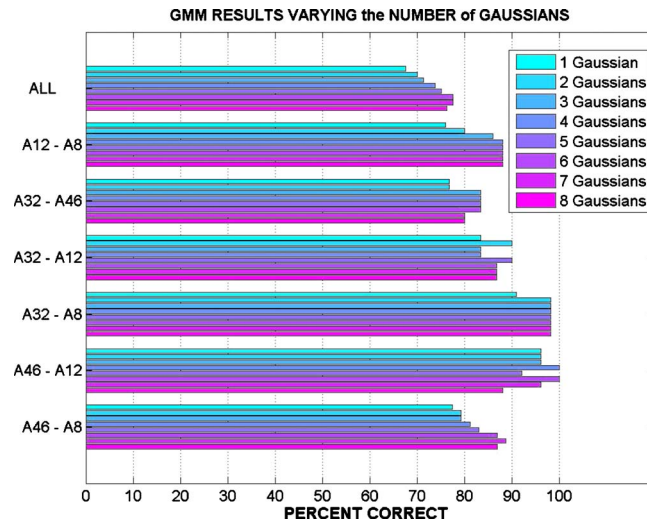


Fig. 2. (Color online) Gaussian mixture model results showing the dependence on the number of Gaussians in the model. Twenty four cepstral coefficients were used as features.

3.4 Discussion

Results are summarized in Fig. 4. The pairwise comparisons with both members from the same matriline (for example A32-A46) were indeed the lowest, indicating greater similarity of calls than pairwise comparisons with members of different matriline. Comparing individual A8 with A32 (different pods) had the greatest discrimination (98% correctly assigned), but the comparisons of A8 with the other two members of the second pod were not as exceptional in discrimination.

Also included in Fig. 4 are the Neural Network (NN) results of Nousek *et al.* (2006). While there was little difference in results between the GMM and HMM calculations, both were



Fig. 3. (Color online) Hidden Markov model results showing the dependence on the number of states in the model. Two Gaussians and 24 cepstral coefficients were used in the calculations

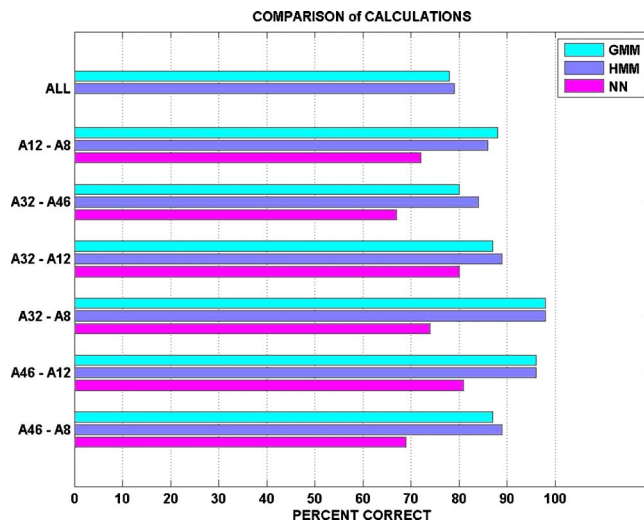


Fig. 4. (Color online) Neural Net (NN) calculation compared to the Gaussian mixture model with 7 Gaussians and the hidden Markov model with two Gaussians and 5 classes. The “ALL” calculation was not carried out with the NN.

better by around 10% or more than the NN calculations. They offer the additional advantage of simpler calculation of the features as was discussed.

4. Conclusion

These results demonstrate that both GMM’s and HMM’s are highly successful in the task of automatic identification of individual killer whales from a sample of known individuals. This is particularly impressive since it is doubtful that humans could listen to these four groups of sounds and then identify the appropriate group for an unknown sound. This method shows promise for tracking trajectories of individual killer whales.

Acknowledgment

We are very grateful to Patrick Miller for the killer whale sounds which made this study possible.

¹<http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>

Adi, K., Johnson, M. T., and Osiejuk, T. S. (2010). “Acoustic censusing using automatic vocalization classification and identity recognition,” *J. Acoust. Soc. Am.* **127**, 874–883.

Brown, J. C. (1999). “Computer identification of musical instruments using pattern recognition with cepstral coefficients as features,” *J. Acoust. Soc. Am.* **105**, 1933–1941.

Brown, J. C. (2008). “Mathematics of pulsed vocalizations with application to killer whale biphonation,” *J. Acoust. Soc. Am.* **123**, 2875–2883.

Brown, J. C., Houix, O., and McAdams, S. (2001). “Feature dependence in the automatic identification of musical woodwind instruments,” *J. Acoust. Soc. Am.* **109**, 1064–1072.

Brown, J. C., and Smaragdis, P. (2008). “Automatic classification of vocalizations with Gaussian mixture models and hidden Markov models,” *J. Acoust. Soc. Am.* **123**, 3345.

Brown, J. C., and Smaragdis, P. (2009). “Hidden Markov and Gaussian mixture models for automatic call classification,” *J. Acoust. Soc. Am.* **125**, EL221–EL224.

Buck, J. R., and Tyack, P. L. (1993). “A quantitative measure of similarity for *Tursiops truncatus* signature whistles,” *J. Acoust. Soc. Am.* **94**, 2497–2506.

Deecke, V. B., Ford, J. K. B., and Spong, P. (1999). “Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects,” *J. Acoust. Soc. Am.* **105**, 2499–2507.

Ford, J. K. B. (1987). “A catalogue of underwater calls produced by killer whales (*Orcinus orca*) in British Columbia,” *Can. Data Rep. Fish. Aq. Sci. No.* **633**, 1–165.

Miller, P. J. O., and Bain, D. E. (2000). “Within-pod variation in the sound production of a pod of killer whales,

Orcinus orca,” *Anim. Behav.* **60**, 617–628.

Miller, P. J. O., and Tyack, P. L. (1998). “A small towed beamforming array to identify vocalizing resident killer whales (*Orcinus orca*) concurrent with focal behavioral observations,” *Deep-Sea Res., Part II* **45**, 1389–1405.

Nousek, A. E. (2004). “The influence of social structure on vocal signatures in group-living resident killer whales (*Orcinus orca*),” MS thesis, University of St. Andrews, St. Andrews, Fife, Scotland.

Nousek, A. E., Slater, P. J. B., Wang, C., and Miller, P. J. O. (2006). “The influence of social structure on vocal signatures in northern resident killer whales (*Orcinus orca*),” *Biol. Lett.* **2**, 481–484.

Oppenheim, A. V., and Schaffer, R. W. (1975). *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Rabiner, L. R., and Juang, B. H. (1993). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).

Rabiner, L. R., and Schaffer, R. W. (1978). *Digital Processing of Speech Signals* (Prentice-Hall, London).

Reynolds, D. A., and Rose, R. C. (1995). “Robust text-independent speaker identification using Gaussian mixture speaker models,” *IEEE Trans. Speech Audio Process.* **3**, 72–83.