Singing maps: Classification of whalesong units using a self-organizing feature mapping algorithm

Ashley Walker, Robert Fisher and Nicholas Mitsakakis
Department of Artificial Intelligence, University of Edinburgh, 5 Forrest Hill, Scotland, EH1 2QL

Singing maps: Classification of whalesong units using a self-organizing feature mapping algorithm

Ashley Walker, Robert Fisher and Nicholas Mitsakakis
Department of Artificial Intelligence, University of Edinburgh, 5 Forrest Hill, Scotland, EH1 2QL

A widespread problem in the study of humpback whale song vocalizations involves evaluating the similarity of song elements within a whale's repertoire, between individuals of a social group, and between social groups separated by time and space. Whilst humpback whale songs demonstrate a remarkable amount of regular high level structure, they are composed of a variety of complex and transient elemental phonological units. Reliable classification of song structure requires robust unit classification — a feature which has made this process difficult to automate. This work presents a fully automated technique for performing multiple-resolution unit classification. In this scheme, units are simultaneous assigned membership to a series of increasingly general acoustic classes such that degrees of song structural similarities (and differences) emerge from analysis of units classified at different resolutions.

PACS numbers:43.80

INTRODUCTION

Humpback whales (*megaptera novaeangliae*) emit long, complex patterned vocalizations, or "songs"[1]. A number of discrete populations of humpback whales exist which, at any point in time, can be characterized by a unique song shared by all singing population members [Tyack 1981; Winn et. al 1981; Payne & Payne 1985]. The songs of the various populations which have been studied all have in common a complex hierarchical structure.

The generally accepted whalesong grammar identifies 5 levels in this structural hierarchy [Payne & McVay 1971; Payne & Payne 1985].

- At the lowest level are primitive phonological *units* — "the shortest (real-time) sounds continuous to human ears".
- *Phrases* are organized groups of units. Typically, a phrase will consist of two distinct units (e.g., a tone '-' and a click '|') each consecutively repeated in variable number times.
- Sequences of one of more similar phrases form a *theme*. Themes may be composed of one of several phrasal patterns — e.g., *static phrases* which contain virtually the same material (e.g., - - | - - | - - |), or *shifting phrases* whose content gradually drifts from one unit type to another (e.g., - - - / / / | | |), etc.
- Sequences of themes (typically 4-10) — arranged in a fairly inflexible order — make up a *song*.
- Singing whales repeat the song that is currently typical of their population in long

[1]Humpback whales also make a variety of social sounds that are heard most often when the whales are interacting in groups [Thompson et. al 1977]. These sounds appear to be subject to different rules from those influencing songs. Moreover, both genders make social sounds whereas almost all observed singing humpbacks have been male [Payne & Payne 1985]. The role that song plays in the lives of humpback whales is unclear. Traditionally it was believed be purely a cultural phenomena — playing a part in courtship analogous to bird song. However, the low-frequency, repetitive, patterned vocalizations of the humpback whale may also/instead be used for environmental sensing [Frazer et. al 1996]. In this paper we refer to these vocalizations as "song" for historical reasons.

(up to many hours) cycles, or *song sessions*.

The duration of songs — particularly the number of phrases per theme and themes per song — varies, even between songs in the same song session sung by the same individual. However, the structure and sequence of this theme grammar is so universally adhered to that the few observed deviations have been labelled as aberrant. Over a series of years, the characteristic song of each humpback population changes extensively and irreversibly (within the confines of the grammar). Just as the songs themselves are highly structured, song evolution appears to be a methodological process.

Techniques for investigating animal phonology have predominantly relied upon a trained observer's subjective impressions of aural cues or visual inspection of sound spectrograms. Manual classification can be a reliable means of categorization [Thorpe 1966; Payne & McVay 1971], but the features upon which such a comparison is based are often illspecified and intuitive.

Automatic classification techniques — such as multivariate statistical and neural network analysis [Clark 1982; Chabot 1988; Potter et. al 1994] introduce higher levels of objectivity at the unit comparison stage. However, many existing automated techniques rely on manual pre-processing (unit detection) and post-processing (grouping of similar units into classes). In the context of whalesong analysis, because we can group units into both *unit-classes* and strings of units into *phrase-*, *theme-* and *song-classes*, automatic techniques which support a hierarchy of classification analyses are needed.

In an effort to fill this niche, we present a fully automatic technique for extracting, comparing and grouping units. This technique provides a novel similarity measure and a multi-resolution clustering space. The multiple scales of classification resolution engender this technique with the robustness and flexibility necessary to support automatic comparisons of high-level song structure.

I. CLASSIFICATION METHODOLOGY

A. Unit Extraction and Representation

Computer based methods of sound classification operate on a numerical representation of a sound's time-frequency spectrum. Narrow-band frequency modulated signals are often represented via one or more contour descriptions extracted from a sonogram [Clark 1982; Dawson & Thorpe 1990; Buck & Tyack 1993; McCowan 1995]. This feature representation provides an efficient and powerful encoding for narrow-band signature sounds such as whistles. However, to represent the variety and complexity of broadband humpback whale sound units, a time-frequency distribution is more appropriate — i.e., it retains more information about each sound and requires less subjective assessment of character relevance.

We represent humpback whale song units using a spectrogram obtained through application of the short-time Fourier Transform (*512*-pt., Hamming filter, *20*% overlap) to samples of recorded whalesong (digitized at *20* kHz). Because human observers perceive whalesong units to vary in duration from approximately *0.1 -10* seconds [Payne & Payne 1985], we define a unit precisely as a continuous sequence of recorded song signal whose energy exceeds a preset threshold[2] for more than *0.1* seconds. Units can vary in length — their boundaries being defined by *0.1* seconds of subthreshold sound.

Extracted *unit spectrogram matrices* consist of *120* time columns — spaced at *0.02* second intervals across *2.47* seconds — with their sound energy centered along the time axis. Matrices containing sound units which are shorter than this maximum duration are zero padded on both ends.

In order to make the representation more computationally efficient, the original *256* frequency linearly spaced channels in the

[2] This threshold was set at the mean plus half of the standard deviation of the maximum amplitude in any frequency channel of transformed song signal segments .

unit spectrogram matrices are compressed into 16 logarithmically spaced frequency rows — the lower 14 of which are used for matching.

As a final step, all matrices are normalized such that the sum of the square of all spectrogram cells equals one, thereby eliminating effects of differing recording or playback levels. The unit bandwidth and duration encoded does not cover the full possible range of whalesong units, but represents the available data accurately. Figure 3 — top rows in blocks 1-3 — shows some examples of extracted, encoded units.

Our encoding is an improvement on the binary spectrogram matrix used by Chabot [1988] in an acoustic taxonomic study of humpback communication sounds. Chabot's units are manually extracted from spectrogram traces and digitized into *16* by *21* matrices. His logarithmically spaced frequency bins span the same range as that we used, but maximum unit duration (*2* seconds) and time resolution (*0.1* second intervals) were lower. Because cell energy was binarized, Chabot additionally coarsely encodes features such as relative intensity, frequency and amplitude modulations as additional binary variables. On the other hand, our representation makes these explicit.

B. Similarity Measure

Techniques for measuring the similarity of sounds encoded in the cells of spectrogram matrices often rely on cross correlation [Clark et. al 1987; Nowicki & Nelson 1990] and related approaches [Chabot 1988]. Here, we define the similarity *Sim(s,w)* of two (normalized) sound signals, *s* and *w*, using an inner product. This similarity measure is robust under high ambient noise — since it enhances common components of the two sounds being compared, but does not encourage negative matches (i.e., matches between cells with little sound energy in both signals).

In computing the similarity of any pair of sounds, we allow units to slide some distance along the time axis to ensure that the similarity measure reflects the best fit match. Whilst centering units along the time axis prevents matching based on sound start characteristics, time shifting further discourages matches between units whose only common feature is sound duration. Moreover, time shifting is essential for the alignment of pulse trains — which if offset by a single time column can alter the match irrevocably [Chabot 1988].

Frequency shifting is not used. Due to the low frequency resolution of our spectrogram matrices, pitch differences remaining after encoding may contain useful information.

Similarity measures used to compare narrow-band frequency modulated sounds, like dolphin whistles, often employ unit expansions and compressions in time and/or frequency [Buck & Tyack 1993; McCowan 1995]. It is not clear in the case of humpback whale song units what improvements might be gained by this. In the case of our limited dataset, we found global warping (via blurring cell contents across adjacent frequency rows and time columns) to be of little value.

We did implement, however, a *local warping* function to facilitate consistent coding of similar units which tend to be sung at subtly different rates (e.g., pulse train repetition rates). Chabot [1988] found it so difficult to keep pulse train coding consist within his binary spectrogram matrix that sampling cells containing silences (i.e., '0's) within obvious pulse trains were re-coded as sound (i.e., represented by a '1'). We do not edit our unit spectrogram encoding, but allow each individual cell in the matrix to stretch a limited amount along the time and frequency axes in order to find the best match during correlation. A penalty function keeps the amount of stretch within reasonable bounds by "charging" the similarity measure for increased amounts of local warping. Each local warp depreciates a cell correlation by:

$$\frac{\alpha}{\alpha + \mathrm{warp}}$$

where $\alpha = 0.5$ and *warp* is the amount of local stretch (measured as a number of cells).

C. Grouping

Comparing pairs of units via this similarity metric allows us to identify groups or classes of acoustically similar sounds. Due to the apparent degree and rapidity of linguistic drift in humpback whalesong phonology, however, the long-term significance of these classes is not easy to assess.

Though humpback whales are not particularly faithful to their units, they do appear to be loyal to a grammar. The fact that more regular, high level song features rely on transient and irregular low level units seems to present a contradiction. However, we argue that this situation presents an opportunity for a new performance criteria to be applied to humpback whalesong unit classifications. Specifically, we propose that the performance of a unit clustering procedure should be based upon how well the underlying classification space supports comparative analysis of humpback song, theme and phrase, as well as units.

A classification space which could provide a robust and convenient organization of units should include the following features:

- *Topological ordering*. Clustering involves mapping *N* groups of neighboring data points in the high dimensional *unit space* to *N* classes in a lower dimensionality *classification space*. If the spatial relationships amongst the classes in the lower dimensional space preserves the inter-group topology of the higher space, then units can naturally have membership in both an acoustic *class* and an acoustic *region* of classification space. This ordering facilitates the multi-resolution unit classification necessary for robust higher level matching.
- *Nonlinear decision borders between classes*. In order to cluster overlapping unit classes into categories possessing maximum internal cohesion and external isolation, nonlinear discrimination borders are necessary. We can expect this sort of adaptive discriminatory behavior from intelligent animals capable of, for example, decoupling the co-occurance of signal and noise.
- *Existence of an explicit cost function*. The existence of a cost function enables ranking of the different possible clusters so as to optimize a particular clustering statistic.
- *Generation of a sound class prototype.* Class prototypes, or averages, facilitate observation of common features of class members.

A clustering process which simultaneously meets all these criteria is Kohonen's Self Organizing Map (SOM) [Kohonen 1988]. The SOM algorithm maps from an input space to a 2-dimensional classification space, or "phonetic map" (in its original application). The phonetic map is not a formant map, nor a kind of principle component graph for phonemes. Instead, "it displays the images of the complete spectra [spectrogram] as points on a plane, the distances of which approximately correspond to the vectoral differences between the original spectra; so this map should be regarded as a similarity graph, the coordinates of which have no explicit interpretation" [Kohonen 1988].

In our implementation, a phonetic, or classification, map is created as a 2-dimensional lattice of artificial neurons or "nodes" each containing a *14x120* prototype spectrogram matrix. Node matrices are initialized by randomly pertubating an average of all the input spectrogram units. Node contents come to represent the characteristic time-frequency distributions of classes in the input space through an iterative training procedure. Specifically, at each epoch in the training cycle, an input s_k is randomly selected and its (unwarped) similarity $Sim(s_k, w_{i,j})$ to each node $w_{i,j}$ in the network is computed. The best match is reinforced by adjusting the winning node's prototype distribution (and the distribution of other nodes within its neighborhood) to better encode the current input distribution:

$$\Delta w_{i,j} = \eta O(f(s_k) - w_{i,j})$$

where O (neighborhood profile) and η (the learning rate) are functions which decrease with time[3]. The function f performs time shifting and local warping to maximally align units with their prototypes before performing an update.

This is a k-means type clustering algorithm with iterative updates. Unlike traditional k-means, this iterative algorithm is not guaranteed to reach a stable local minima (i.e., stable, locally optimal clusters), but decreasing the learning rate and neighborhood size encourages convergence. This algorithm lets us specify a cost function (based on the squared error) which ranks the local minima. Stochastic noise introduced by the input presentation order is thought move the clustering from higher minima toward progressively lower ones (though, again, not necessarily the global one).

II. RESULTS

In this section, we demonstrate the automatic classification technique. These results, obtained for a limited data set, are not intended to yield answers to behavioral questions, but to serve as a proof of the validity and utility of the proposed technique.

A. Unit Extraction and Representation

The harmonic structure of whalesong units is complex — varying from tonal to noise bursts, across a wide range of intensities. Nevertheless, employing only a knowledge of unit durations, inter-unit silence durations and a fairly arbitrary signal-to-noise detection threshold, the simple extraction algorithm detected units with an 88% success rate when compared to manual extractions.

The value of the thresholding parameter caused the automatic extraction process to miss quiet units which human observers could discriminate from noise. On the other hand, the duration parameters were effective: pulse train units present in our data had a repetition rate in excess of *0.1* seconds and therefore were extracted as integral units — as they would be via manual extraction.

B. Similarity Metric

We investigated the similarity function using several examples of four types of units extracted from our data. ("Types" were manually classified for the purposes of this investigation.)

In Table I, the similarity *Sim(s,s)* values — with time shift (plus local warping) — for all possible combinations of internal and external class comparisons are listed. As we only apply local warping during the node update phase of clustering, local warping values are only shown for internal class comparisons.

All similarity scores increase with local warping. As expected, local warping improves matches between units with highly variable rates — e.g., pulse trains — most significantly.

	low tone	**high tone**	**pulse train**	**FM sweep**
low tone	0.668 (0.738)	0.290	0.248	0.360
high tone		0.840 (0.885)	0.251	0.241
pulse train			0.579 (0.673)	0.224
FM sweep				0.706 (0.760)

TABLE I. Average similarity values for internal and external unit class comparisons. (Parenthesized values are computed with local warping.)

C. Grouping

[3] Satisfactory values for the learning rate (η = *0.7 down to 0.0*) and neighborhood size (*N = 6 down to 0*), neighborhood profile (i.e., "Mexican hat function") and decay rates (linearly with time) were discovered through experimentation. In determining these parameters, trials which resulted in "successful" maps showed the greatest degrees of node/class internal cohesion and external isolation. At least one hundred presentations of our data (504 automatically extracted units from a single song file) was required in order for the network to learn the clusters. Training takes approximately 5 hours on a Sun Workstation (110MHz), Sparc 5 architecture. Software written in C.

Figure 1 shows the results of mapping a complete song onto a classification network. Each of the 49 basic acoustic classes is represented via a class prototype spectrogram matrix. (Lighter regions denote louder sound.) It is immediately apparent that the algorithm learned to distinguish the familiar time-frequency structures which govern manual clustering. There are tones: long duration, high frequency (top right), short duration medium frequency (bottom left) and very low frequency, very short tones with higher harmonics (bottom right). Frequency sweeps start at the top left of the map and change their profile gradually as one moves obliquely (down and to the right).

Because of the topology preserving nature of the SOM, acoustically similar spectrogram prototypes are grouped into classification neighborhoods, or regions. Within these regions, the network distinguishes between similar time-frequency profiles based upon unit duration and amplitude modulation characteristics.

Notice also that enhanced spectral representation of the units (i.e., encoding relative power distribution as well as the single fundamental frequencies) allows greater discrimination between units. For example, unit (*4,1*) has power concentrated over a shorter, but broader spectral peak than (*7,1*), whose power is spread over a longer time. It is unclear if there is any behavioural significance to these distinctions.

The size of the network determines how well we can detect these subtle differences between units. As our objective is not to perform a taxonomic study of the humpback whale song repertoire, we employ a qualitative criterion to chose the network size.

For example, if we defined a small classification network, the algorithm produces general acoustic classes — each representing a broad range of acoustic features. Songs represented by this coarse-scale encoding would contain a few heavily repeated units and sequences of units (i.e., phrases and themes). In this scenario, the prototype class spectrograms could not represent any individual unit well, and the inferred song structure is unrealistically simple.

A very large network would have the opposite effect. Individual units would be well represented by their prototype (in fact, each could *be* a prototype in the case where the network size matched the data size), but few regular patterns would exist in the sequences of classified units: the song structure would appear overly complex. This occurs because the "repeating units" which make up what we know as phrases and themes are not numerically identical sounds but vary subtly in duration, frequency range, sweep rate, etc.

For this reason, the network size is chosen so as to compress the data into a manageable number of classes (e.g., a *10:1* reduction in number of units:classes), while allowing enough acoustic diversity that prototypes correlate well (e.g., $Sim(s,w) > 0.5$) with the original sound units they represent.

The *7 x 7* network shown in Figure 1 achieves a nice compromise. With one exception, each of the nodes in the network best represents one or more units in the input data. Twelve percent of the node classes are used repeatedly (i.e., they each encode *>3%* of the data) and *25%* specialize in less than *1%* (each) of the data. (Node *use* scores are shown above each class prototype spectrogram in Figure 1.) By assigning units to a single node class we can perceive subtle differences between units.

However, because of the topological ordering of the classification space, units are also well represented by the several class prototypes which surround their primary class assignment. Figure 2 shows the similarity profile of a typical unit — where the similarity values of a unit and all 49 network nodes are sorted and displayed in order. Similarity values are high between a unit and its *1-3* nearest neighboring nodes, and fall off steeply outwith this region.

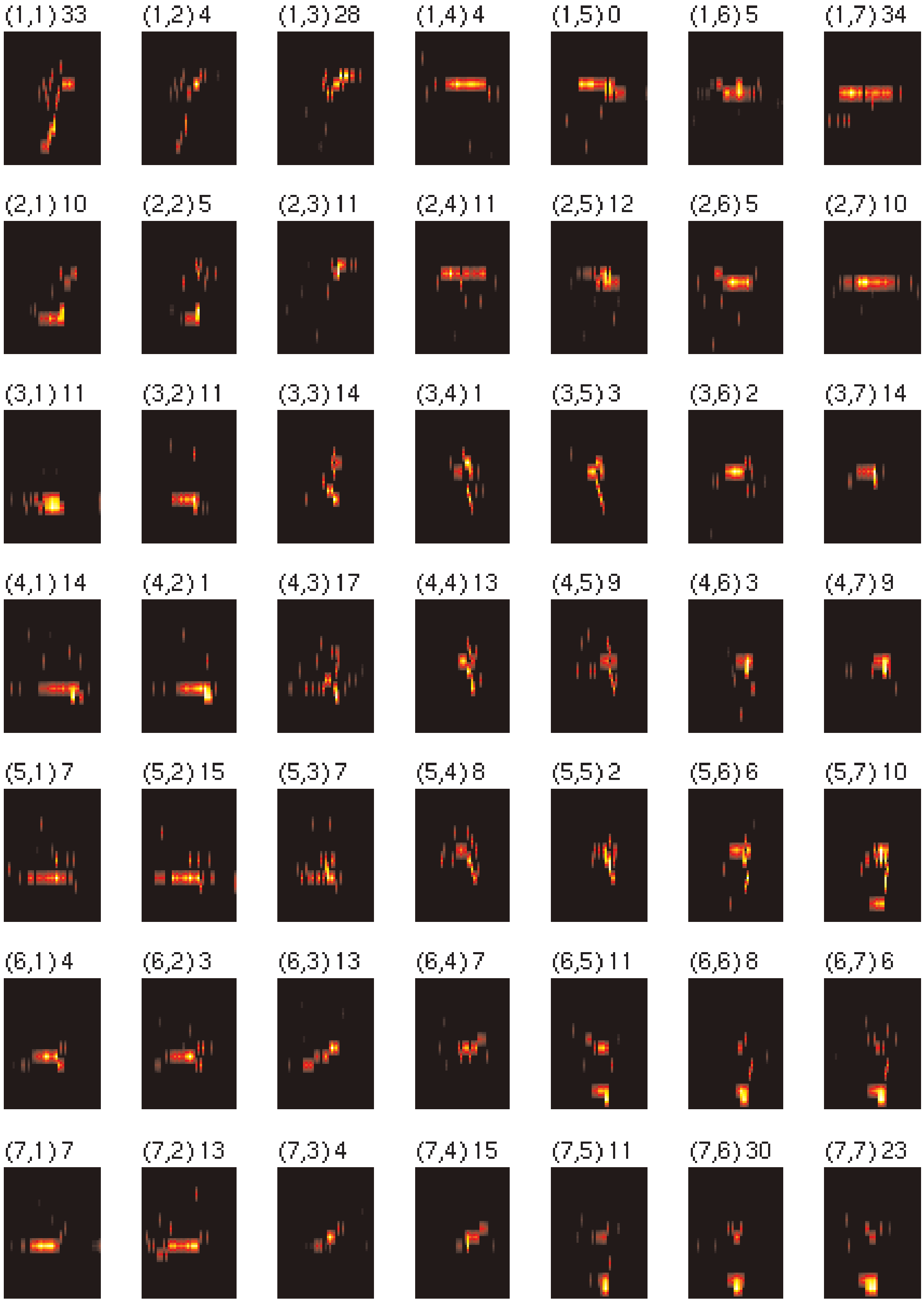

FIG 1. Phonetic map.
49 acoustic class prototype spectrogram matrices. (Coordinates) Usage.

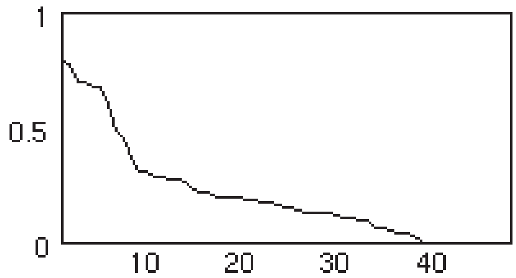

FIG. 2. Similarity profile for a typical unit. (Axes are node vs. similarity score.)

By assigning units to a class region, we can ignore subtle differences between units which may play the same role in the song structure.

III. Discussion

A. Unit Extraction and Representation

The automatic unit extraction mechanism presented here is a simple, but powerful tool for rapidly parsing large amounts of data. If the algorithm employed more information about the physiological boundaries of a whale's song production mechanism (e.g., frequency range), we would expect better agreement between automatic and manual extractions. Moreover, for a particular recording environment, filters could be created to detect interference with characteristic sound patterns, e.g., propeller noise.

Assessment of the quality of the representation used to encode extracted units is beyond the scope of this investigation. Until more is known about humpback whale audition, conclusions about representational issues will be limited to comparisons between published schemes. However, it is generally recognized that a constant Q spectrographic technique like a wavelet transform may be necessary to obtain biologically realistic resolution in both the time and frequency domains simultaneously [Cohen 1989; Groutage et. al 1992; Potter et. al 1994]. The traditional spectrogram is used in our studies because of the ease with which results can be compared with other work in the field.

By contrast to most other approaches of which we have knowledge, our spectrogram matrix representation uses all parts of the recorded sound — aside from absolute intensities, which are lost in normalization. Filtering lower intensity sounds, harmonics or sidebands introduces subjectivity (e.g., the fundamental frequency is often difficult to isolate) and is difficult to justify given the small amount that is known about the humpback auditory system [Campbell 1963]. Masking techniques omit many of the small acoustic details which may be an artifact of natural variation, but may likewise be of behavioral significance.

We are uncertain as to whether the resolution used in this representation is sufficiently precise to match the sensitivity of the humpback auditory system. In particular, it would be interesting to know whether our representation has enough resolution along the time axis to encode the pulse train repetition rates (and rates of change) perceivable to the humpback whale. To our knowledge, the ability of humpback whales to discriminate this variable has not been studied. Odontocetes, however, are extremely sensitive to time intervals between their echolocation clicks [Thomas & Kuechle 1982].

B. Similarity Measure

The similarity measure proposed here uses more domain knowledge than any other technique we have seen for comparing whalesong units. Due to time shifting, it is relatively insensitive to the absolute temporal alignment of sounds and their duration. The latter feature is used only in fine-detail discriminations between units with similar time-frequency profiles.

Frequency shifting might be employed in future work should it become known that pitch differences carried no information — e.g., were an artifact of a singer's age.

No mechanism for performing global expansion or compression of units is used here. Extracted feature representations (e.g., frequency sweep contours) are more naturally globally warped than are spectrograms.

However, it is possible, through sub and super-sampling methods, for this type of warping to be applied to spectrograms. There are some behavioural studies for which global warping might be useful. Payne & Payne [1985] report that song themes sung in adjacent years are sung at different rates (with the later year's song being rendered more slowly). This comes about through expansions of unit (and inter-unit silence) durations, and through a process of splitting units which were once contiguous into two distinct sounds.

A mechanism for performing local warping is proposed here and it is shown to encourage better matching amongst units which have an intrinsic rate (pulse trains). The penalty function which limits the amount of local warping could be optimized in a later study.

C. Grouping

Unit classification for the purpose of automatic analysis of high level song structure is not simply a unit labelling process. Multiple resolution techniques for clustering allow us to explore classification generalizations to arrive at the best scale for differentiating song units.

The utility of SOM multi-resolution representation can be seen in the example presented in Figure 3. There, a series of consecutive units which form what we consider to be a phrase are shown along the top row of each phrase block. (Three consecutive repetitions — blocks 1-3 — of this phrase are shown.) This phrase consists of two types of units (where a "type" is a categorization determined by our observation): a variable number of low tones with higher harmonics, followed by a few higher frequency, purer tones.

The similarity scores of the best three matches for each unit are shown in the middle row of each block of Figure 3. The first sound unit is most similar (*Sim=0.6722*) to node (*6,5*). We could call that node its "class", however, doing so would put it in a different class from the next sound unit (which maps to node (*7,7*), *Sim=0.7095*). The first unit has more energy in the higher harmonic, however, these two units come from adjacent positions in the song sound file and their acoustic differences may not be behaviorally significant.

For the purposes of bringing out the structure in a phrase, we classify units by class *region*. For example, we could define a region via a center node (e.g., (*6,5*)) and a radius (e.g., *+/-2* nodes). Classifying units via this region-based scheme would allow us to automatically identify the phrasal structure — i.e., the (5-7) low tones with higher harmonics, followed by a (1-3) higher, longer tones — apparent to human observers.

The final row of each phrase block in Figure 3 shows the pattern of similarity values that each unit induces across the network. (Larger similarity values are encoded with brighter colors.) Tracking patterns of similarity across the network (in response to a string of consecutive song units) provides another convenient window through which to "see" high level patterns in the song emerge.

In the case of the phrase shown in Figure 3, the regions of high similarity values clearly distinguish the units at the beginning of the phrase from those at the end. Moreover, these presentations make explicit some behavioural points of interest. The phrase seems to conclude with 2 different types of tones. The significance of this is unclear. Also, note that upon each successive repetition of the phrase, the highly repeated units become more regular (i.e., they map increasingly to node (*6,7*) and vary over a smaller radius (*+/-1*)). We observed this phenomena in other phrases from this song file.

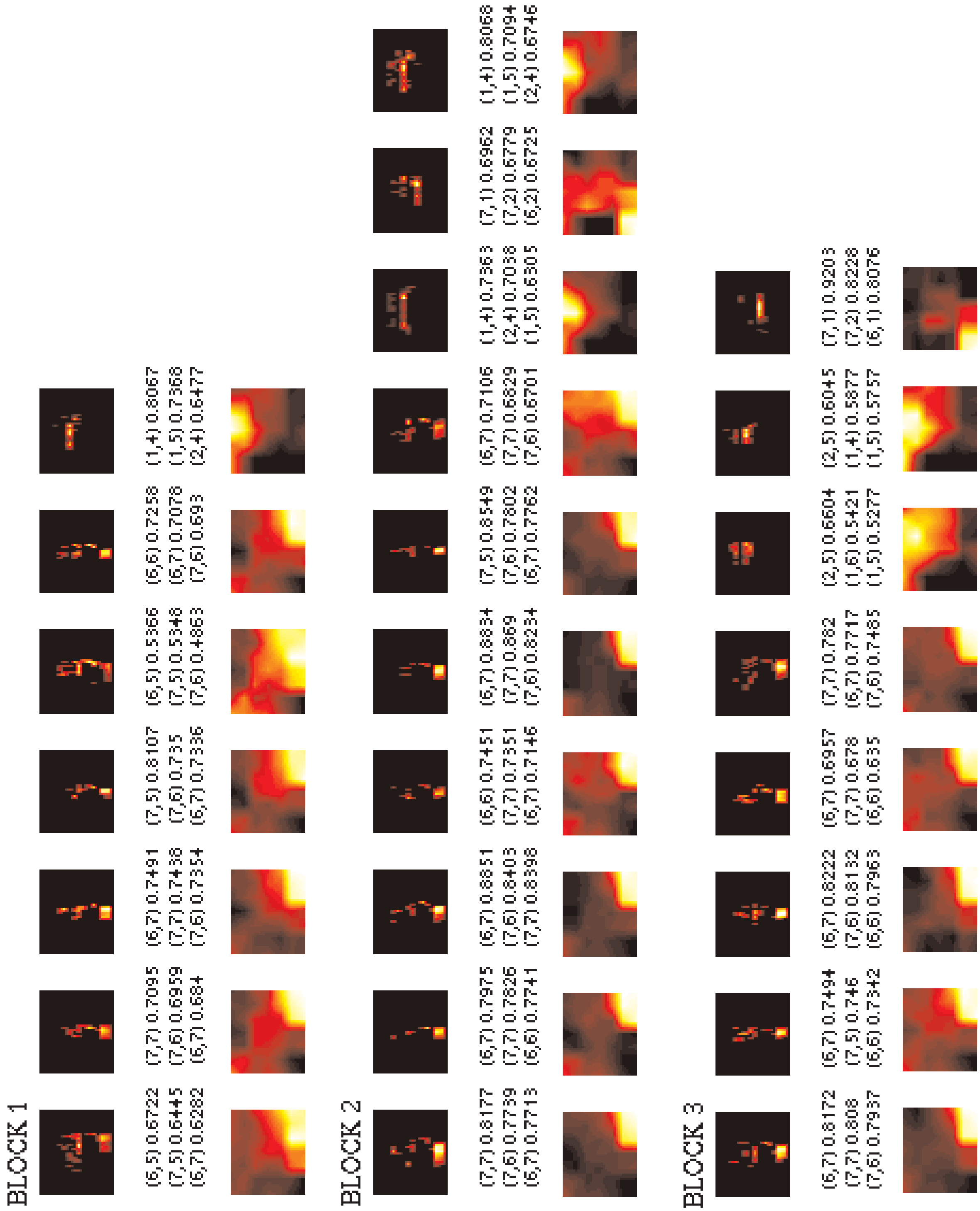

FIG. 3 Song phrase.
Top row, BLOCKS 1-3: Unit spectrogram matrices.
Middle row: Best three class assignments. (Node position - see FIG 1.) Similarity.
Bottom row: Similarity scores across the network. White denotes high correlation, black low.

IV. CONCLUSIONS

Over the past decade, humpback whalesong research has begun to address numerous issues regarding song structure:

- rate of song development within populations,
- song learning mechanisms,
- cross population comparisons of songs, and
- characteristics of aberrant songs.

Automated techniques for classifying high level song features facilitate direct comparison of answers to these (and other) problems. Automated classification tools allow us to parse more data (more uniformly) and, most importantly, make classification criteria explicit.

Successful automated classification mechanisms will rely on the robust and flexible unit classification at which human observers excel. The technique presented here outlines a strategy for performing this sort of classification. In particular we have demonstrated a simple and reliable method for automatically extracting units, an improved measure for comparing them and the first classification space of which we know that creates multiple hierarchical levels of unit classification.

ACKNOWLEDGMENTS

We would like to thank Fabrizio Carrera and Martin Westhead for digitizing (and listening to) numerous audio recordings.

REFERENCES

Buck, J. & P. Tyack (1993). "A quantitative measure of similarity for *Tursiops truncatus* signature whistles", *J. Acoustic Soc. Am.*, **94**, 2496-2506.

Campbell, R. (1963) "Frequency discrimination of pulsed tones", *J. Acoustic Soc. Am.*, **35**, 1193-1200.

Chabot, D. (1988). "A quantitative technique to compare and classify humpback whale (*Megaptera novaeangliae*) sounds", *Ethology*, **77**, 89-102.

Clark, C. (1982). "The acoustic repertoire of the southern right whale, a quantitative analysis", *Anim. Behav.*, **30**, 1060-1071.

Clark, C., P. Marler & K. Beeman (1987). "Quantitative analysis of animal vocal phonology: an application to swamp sparrow song", *Ethology*, **76**, 101-115.

Cohen, L. (1989). "Time-Frequency distributions - a review", *Proc. IEEE*, **77**, 941-981.

Dawson, S. & C. Thorpe (1990). "A quantitative analysis of the sounds of Hector's Dolphin", *Ethology*, **86**, 131-145.

Frazer, N., E. Mercado & A. Tolstoy (1996). "Understanding humpback whale sonar: A physicist's view", *J. Acoustic Soc. Am.*, **100** (4), 2644.

Groutage, D., J. Schempp & L. Cohen (1994). "Characterization and analysis of marine mammal sounds using time-frequency and time-prony techniques", *Ocean*, **1**, 253-258.

McCowan, B. (1995). "A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, *Tursiops truncatus*)", *Ethology*, **100**, 177-193.

Mitsakakis, N. (1996). "Classification of humpback whalesong units using a self organizing feature map", *MSc Thesis*, Department of Artificial Intelligence, University of Edinburgh.

Nowicki, S. & D. Nelson (1990). "Defining natural categories in acoustic signals: comparison of three methods applied to chick-a-dee' call notes", *Ethology*, **86**, 89-101.

Payne, R. & S. McVay (1971). "Sounds of humpback whales", *Science*, **173**, 585-597.

Payne, K. & R. Payne (1985). "Large scale changes over 19 years in songs of humpback whales in Bermuda", *Zeitschrift fur Tierpsychologie*, **68**, 89-114.

Potter, J., D. Mellinger & C. Clark (1994). "Marine mammal call discrimination using artificial neural networks", *J. Acoustic Soc. Am.*, **96**, 1255-1262.

Thomas, J. & V. Kuechle (1982). "Quantitative analysis of Weddell seal (Leptonychotes weddelli) underwater vocalizations at McMurdo Sound, Antarctica", *J. Acoustic Soc. Am.*, **72**, 1730-1738.

Thompson, P., W. Cummings, & S. Kennison (1977). "Sound production of humpback whales, *Megaptera novaeangliae*, in Alaskan waters", *J. Acoustic Soc. Am.*, **62**, 89.

Thorpe, W. (1966). "Ritualization in ontogeny. II. Ritualization in the individual development of bird song", *Philos. Trans. R . Soc. London Ser. B*, **251**, 351-358.

Tyack, P., (1981). "Interactions between singing Hawaiian humpback whales and conspecifics nearby", *Behav. Ecol. Sociobiol.*, **8**, 105-116.

Winn, H., T. Thompson, W. Cummings, J. Hain, J. Hudnall, H. Hays, & W. Steiner (1981) "Songs of the humpback whale - population comparisons", *Behav. Ecol. Sociobiol.*, **8**, 41-46.