INVESTIGATION OF AN ARTIFICIAL NEURAL NETWORK FOR RECOGNITION OF SIMULATED DOLPHIN WHISTLES

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE
IN
INFORMATION AND COMPUTER SCIENCES

May 2001

By
Timothy F. R. Burgess

Thesis Committee:

Kazuo Sugihara, Chairperson
Philip Johnson
Martha Crosby

We certify that we have read this thesis and that, in our opinion, it is satisfactory in scope and quality as a thesis for the degree of Master of Science in Information and Computer Science.

THESIS COMMITTEE

____
Chairperson

____

____

ABSTRACT

It is known that dolphins are capable of understanding 200 ‘word’ vocabularies with sentence complexity of three or more ‘words’, where words consist of audio tones or hand gestures. An automated recognition method of ‘words’ where a word is a defined whistle, within a predetermined acceptable degree of variance, could allow ‘words’ to be both easily reproducible by dolphins and identifiable by humans. We investigate a neural network to attempt to distinguish four artificially generated whistles from themselves and from common underwater environmental noises, where a whistle consists of four variations of a fundamental whistle style. We play these whistle variations into the dolphins normal tank environment and then record from a separate tank hydrophone. This results in slight differences for each whistle variation’s spectrogram, the complete collection of which we use to form the neural network training set. For a single whistle variation, the neural network demonstrates strong output node values (greater than 0.9 on a scale of 0 to 1). However, for combinations of ‘words’ (i.e. More than one), the network exhibits poor training performance and an inability to distinguish between words. To validate this, we used a test

set of 41 examples, of which only 22 were correctly classified. This result suggests that an appropriately trained backpropagation neural network using spectrographic analysis as inputs is a viable means for a very specific whistle recognition, however a large degree of whistle variation will dramatically lower the performance of the network, past that required for acceptable recognition.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

SOFAR - SOund Fixing And Ranging

SIMD - Single Instruction Multiple Data

IDL - Interactive Data Language

TSSE - Total Sum Squared Error

1. Sound and Dolphins

1.1 Introduction

It has been demonstrated that dolphins are capable of understanding 200 ‘word’ vocabularies with sentence complexity of three or more ‘words’, where words consist of audio tones or hand gestures. An automated recognition method of defined whistles where dolphin training methods have given definite meaning to these whistles would be a valuable step towards a consensus language. Unlike languages defined in tones or hand signals, a whistle language is in a form which is naturally reproducible by dolphins and should allow meaning defined by whistles to be easily conveyed between dolphin and human researchers.

An ability to cater for a degree of variance is required in the recognition system as sound characteristics through water, harmonic aspects and biologically produced whistle variations occur. This study investigates the use of a neural network method for recognition. Given this architecture’s well recognized pattern classification ability through self-adaptive training and ability to generalize, it would appear to be a viable method to recognize particular whistles with common characteristics.

1.2 Sound underwater

Sound is a form of energy transmitted by rapid pressure changes in an elastic medium. Sound intensity decreases as it travels through seawater because of spreading, scattering and absorption. The intensity loss due to spreading is proportional to the square of the distance from the sound source. Scattering occurs as sound bounces off bubbles, suspended particles, organisms, the surface , the bottom or off other objects. Absorption of sound is proportional to the square of the frequency of the sound I.e. higher frequencies are absorbed more quickly.

The speed of sound in ocean seawater (of 35% salinity) is about 1500 meters per second (3345 miles per hour), almost five times the speed of sound in air. Temperature and pressure affect the velocity of underwater sound however generally speaking, sound travels faster at the warm ocean surface than it does in deeper, cooler depths.

Experimental U.S. Navy depth charges detonated in the minimum velocity layer (or SOFAR layer) in the Pacific have been heard 3,680 kilometers (2300 miles) away from the explosion [6].

1.3 Use of sound by cetaceans

Because sound travels so efficiently through water, many marine animals use sound rather than light to "see" in the ocean. Humans however, evolved our sound senses in air, so we cannot effectively use sound for directional purposes underwater [1],[6]. Our ears would effectively have to be five times further apart. Having evolved in the ocean however, sound is thus the primary sense for cetaceans. Cetacean species use sound to navigate, feed and communicate with each other. Some sub-orders such as humpbacks (mysteceti) are particularly melodious and the complex syntax of their 'songs' suggest rich interaction among individuals.

Bottlenosed dolphins (tursiops truncatus) are members of the cetacean suborder odonceti (toothed whales). Toothed whales have a high brain-weight to body weight ratio and much of their brain tissue is involved in formulating and receiving the sounds on which they depend for feeding and socializing. Toothed whales also use sound to search for prey using echolocation, the biological equivalent of sonar. They generate sharp clicks and other sounds that bounce off prey species and then return to be recognized. Reflected sound is also used to build a picture of the animal's environment and to avoid hitting obstacles while swimming at high speed.

Sounds are produced by dolphins in the range of 5kHz to 150kHz. Echolocation and communication means fall into three general categories. Trains of broad spectrum clicks (or codes), pure-tones that are often frequency modulated whistles and burst-pulse sounds also consisting of clicks but within a different envelope. Some odonceti members (such as sperm whales) only emit clicks. [7]

Current theory is that clicks are used largely for social interaction. There is also strong evidence to suggest that a component of dolphin whistles is an individual's signature characteristic [7]. Present signature whistle studies strongly support the hypothesis that signature whistles are used to maintain group cohesion [12] and that they contain key characteristics for induvidual recognition [23].

When considering whistle classification and recognition, it is also important to consider that dolphins may regard whistles with similar frequency contours in different bands as of the one type [19].

1.4 Animal communication

Although there is by no means a broad scientific consensus on the capability of non-humans to use 'language' (the term 'language' shares the same indefinable qualities as the term 'intelligent'), there have been some impressive gains made in the field of animal cognition. In addition to developments with non-human primates, other species have been found to have communicative abilities previously unknown. These studies indicate a high likelihood exists for human to non-human communication in the future.

Savage-Rumbaugh et al [22] relate the history of Kanzi, a bonobo ape that at 6 months of age was deemed to be too young to undergo the training with the lexigrams that allowed common chimpanzees to communicate. These lexigrams are arbitrary symbols arranged on a board that apes can point to that stand in for words such as banana, look, goodbye and so on. Whilst his mother's performance was disappointing, the research team were surprised to discover that after being separated from his mother at age 2 and a half, Kanzi seemed to be adept at using the lexigrams without having been explicitly taught. Furthermore, and more importantly, Kanzi seems to be able to understand an impressively broad subset of spoken English. His cognitive ability is approximately that of a three and a half year old child. Language has always been regarded as a biological trait unique to humans that failed to evolve with other species. However it appears that some language capability is dormant in other species and in early years of development, it can surface with the appropriate environment.

In a much earlier study, Herman [8], demonstrated the ability of two bottlenosed dolphins to understand sentences in two artificial grammatical command languages. Both semantic and syntactic features of the two languages were shown to be understood. Using words that represented agents, objects, object modifiers and actions, the dolphins showed comprehension for all of the sentence forms and meanings that could be generated by the lexicon and set of syntactic rules.

The two languages consisted of computer generated sounds and a gesture-based visual language. That the dolphins were able to cope with different language mediums with roughly equal success implied that the cognitive skills underlying comprehension competency in the dolphin are very general and not specialized with respect to either the auditory or visual modality.

Cognitive investigations of species other than primate and aquatic species has also produced evidence of strong cognitive capability. Over the last several years, Pepperburg [17] has shown that the African Grey parrot can master tasks once thought to be beyond the capacity of all but humans or certain nonhuman primates. 'Alex' has demonstrated the ability to produce and comprehend 40 object labels (paper, key, nut, etc.), several colors and five shapes. He can combine labels to reference over 100 different objects with an 80% accuracy. This is further validation that the potential exists for human to non-human communication that was unthought of a few decades ago.

Although these findings have important implications for teaching linguistic skills to retarded humans as well as revealing the evolutionary and biological basis for human language, it is important to temper ideological concerns with realism and refrain from over interpreting the results. Many linguists continue to cling to a view that does not allow them to accept the results of the Rumbaughs' and Pepperburg's research. However, an increasing number of cognitive psychologists are applauding these reports of advanced cognitive and rudimentary linguistic abilities.

1.5 Artificial Whistles

We use four artificial whistles (and variants of) that have been generated by Earthtrust researcher, Dr Ken Marten. Based at Sea Life Park, Oahu, Hawaii, Earthtrust is presently working with three female bottlenosed dolphins (Puna, Laukani and Nehoa) in an ongoing program to associate meaning with particular whistles. Observational learning techniques (as used with Kanzi and Alex) are being employed to associate whistles with particular objects. Associating meaning to a dolphin reproducible whistle is effectively forming building blocks of a consensus language between humans and dolphins. This seems feasible given previous work [8] in which dolphins demonstrated understanding of sentence complexity of three words.

Figure 1.1 - Dolphin at Earthtrust lab tank

Dolphins can reproduce artificial whistles as they have demonstrated mimicry of other artificial whistles previously [4]. It is important to note however that an animal simply mimicking sounds is no demonstration of capable intelligence. Construction of meaningful combinations of such words(or whistles) demonstrate such.

A whistle recognizer is an important component of such a language as it enables the human ear to not only determine the presence of a sound but also identify a particular word. The average human hearing is sensitive to 15kHz. A person with very good hearing is sensitive to 20kHz. Given that dolphins produce sounds from 5kHz to 150kHz, humans can only hear sounds at the very low end of the range usable by dolphins.

The artificial whistles we generated were in the frequency spectrum of 5kHz to 11kHz. This allows for the presence of any whistle to be heard by a human ear and possibly also for the frequency changes (heard as a shift in pitch) to allow identification of the whistle. Thus, this frequency range is effectively that which evidence suggests overlaps between upper human capability and lower dolphin capability. Given dolphins are capable of much higher frequency hearing, it may be possible to provide an extended space for more whistle variation by using higher frequencies. For the purposes of validating recognition however, an 11KHz peak is adequate for recognition purposes.

The 11kHz peak was also chosen due to the characteristics of digital sound sampling and future requirements for real-time recognition. In order to properly represent a signal, the sampling rate must be twice that of the reproducible frequency range (Nyquist’s Theorem). Thus an

11kHz signal must be recorded at, at least 22,000 samples per second (22kHz). 22,000 samples is a considerable amount of information to pass to a neural network for classification of a single second of sound so 22kHz was chosen in the interest of reasonable computational performance.

2. Artificial Neural Networks

2.1 Capability and characteristics

The categorization of patterns into separate classes is crucial to the study of animal behaviour. Traditionally animal behaviour researchers have classified behaviour patterns through careful observation by eye. However manual classification has been increasingly replaced by computer methods, understandable due to the likely tedium!

Computerized methods of dolphin whistle analysis include comparison of cross-correlation coefficients using hierarchical cluster analysis and comparison of the average difference in frequency along whistle contours [3]. However an examination of these methods showed they perform less successfully than with common manual comparative techniques [13]. More recent advanced and complex methods which use generalized correlation functions to estimate dynamic models [14] have provided stronger performance.

Neural networks are capable classifiers of patterns and have been successfully used to classify visual and audio patterns. Neural network uses of sound range from examination of components by NASA [26] to speech analysis [18] to underwater acoustic signal processing [15],[5],[25].

Neural networks have been used in dolphin bioacoustics but have primarily focused upon aspects of target discrimination [2], [21] in order to determine the importance of echo features and biological aspects. Most recently, other studies have used neural networks to classify naturally occurring dolphin whistles [11].

Artificial neural networks (ANN's) are well-suited to tasks such as whistle recognition as they are powerful adaptive classifiers. Conceptually based on the brain's structure, they consist of an interconnected assembly of simple processing elements, known as nodes (or neurons) whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-node connection weights. Initially randomized, these weights are adapted from a set of training patterns, which varies weights to minimize error on each training pattern.

Artificial neural networks are typically implemented in software (due to flexibility) and are a network of neurons (or nodes) of typically three to four layers (see Figure 2.1). Each neuron in a layer is connected to each other neuron in neighbouring layers. Each of these connections has an associated weight. A neuron has an internal sigmoidal function which receives it's inputs from the neuron connections and a single output value (see Figure 2.2).

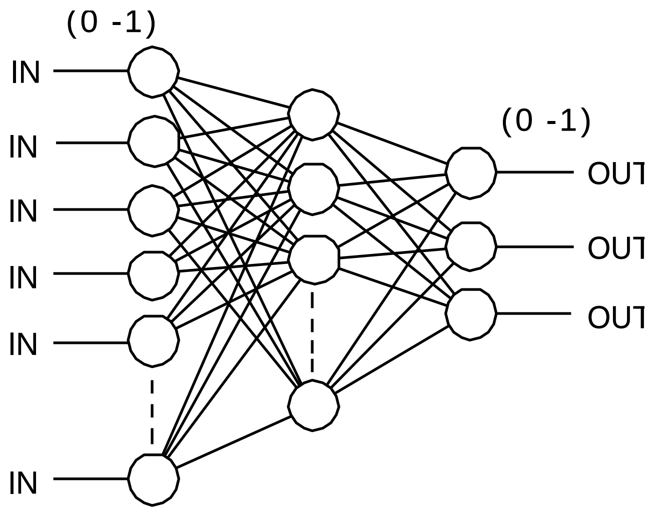

Figure 2.1 - Example Neural Network Architecture

Information (as normalized values between zero and one) is passed into the first network layer, commonly called the input layer. Each layer of neurons from then on receives a value from each input neuron. This value is multiplied (weighted) by the connection weight and then the value is summed and passed through a neuron's sigmoidal function. The output is then passed on to each neuron in the next layer until the output neurons are reached. The backpropagation method is a particularly powerful way in which the degree of error between the required output and the actual output is used to decrementally adjust the network connection weights from the output layer to the input layer. Thus, networks are typically designed and undergo a training period in which an initial series of random weights is adapted to better categorize the input vectors being trained upon.

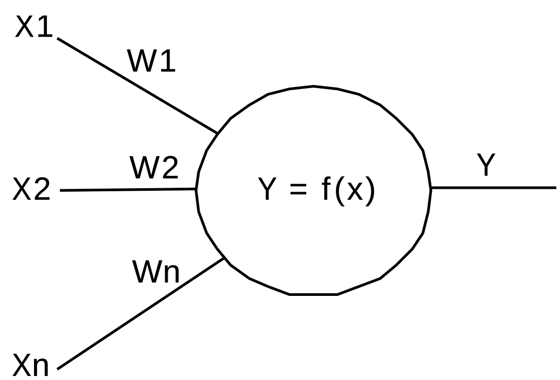

Figure 2.2 - Individual Node with weights and inputs

Compared to traditional software applications written in procedural methods, neural networks have quite different properties, namely:

- the style of processing is completely different - it is more akin to signal processing than symbol processing. The combining of signals and producing of new ones is to be contrasted with the execution of instructions stored in memory

- information is stored in a set of weights rather than the program itself. The weights are supposed to adapt as the network is shown examples from a training set

- ANN's are robust in the presence of noise: small changes in an input signal will not drastically affect a nodes output

- ANN's are robust in the presence of partial node failure: a change in a weight may only affect the output for a few of the possible input patterns

- High level concepts are represented as a pattern of activity across many nodes rather than as contents of a small portion of memory

- ANN's can deal with previously unseen patterns and can generalise from the training set

- ANN's are good at 'perceptual' tasks and associative recall, tasks that a symbolic approach has difficulties with.

The disadvantages of artificial neural networks is that as a self-learning process where the decision criteria are stored as a set of weights, it can be difficult to extract these criteria from the network. Thus the neural network acts as a ‘black box’. Decision criteria can be extracted but it is often a non-trivial exercise to write symbolic procedural code from such.

Large neural networks have a very large number of nodes and thus inherently require a great number of dot-product computations. On a single processor architecture, these calculations can take time to execute sequentially. Multiprocessor architectures however are better suited to neural networks as node calculations are inherently independent and can take strong advantage of parallel CPU’s.

3. Methodology

3.1 Software and hardware

Our primary intention was to confirm that a neural network can consistently perform as a whistle recognizer with a secondary consideration of data rates in order to allow close to realtime performance at a future date. We developed a backpropagation neural network in the C programming language for high performance. This software takes advantage of SIMD (parallel instruction execution) of the Motorola G4 CPU. The software is also multithreaded (capable of parallel operations on a multiprocessor). There was some initial investigation into the possibility of using raw sound data however we chose to accept a more classical approach of deriving spectrograms by Fast Fourier Transforms (FFT) for two reasons.

Firstly, the artificial whistles are visually distinguishable from each other in a spectrographic representation. Thus any data loss caused by using a spectrogram should not be of concern as enough feature details should remain for a neural network to identify whistles.

Secondly, as a frequency representation, a spectrogram allows easy subtraction of frequency components we are not interested in. For our requirements, these are frequencies of 4kHz and below. Our software simply prepares data for neural network training by discarding unwanted frequencies so they are not used in training.

In order to produce spectrograms we chose Research Systems Inc (http://www.rsinc.com) Interactive Data Language (IDL). In addition to having a large number of pre-built functions such as FFT's, IDL provides a number of visual methods for image representation as well as support for the platform which our neural network code has been optimized for (namely Mac OS on G4 computers). As importantly, the IDL code can be further customized as our needs dictate (i.e. different FFT attributes, menus and/or socket communications to the neural network runtime).

The IDL code takes sample windows from the sound files at stepped increments, allowing variation of the sample window and increment size. Each of these windows has an FFT transform performed on it, which provides a frequency intensity representation over the zero to 11khz range of 128 values (for a window size of 256 samples, 128 result values would result). These values then have a logarithmic function performed to transform them to a decibel value.

In order to provide a fixed number of neural network inputs (as the neural network architecture needs to be fixed for training purposes), we use a 3/4 second duration, which provides 164 windows (using 256 samples per window). The sample range of 4kHz to 11kHz provides 102 frequency bins. This results in a total of 16,640 input nodes in the network design.

Using a ceiling value of 80 decibels, as all whistles are quieter than this, we then normalize the 16,640 input values to between zero and one, as required for the purposes of a feed-forward backpropagation network.

As neural networks begin with randomized weight values, initial training session values invariably differ with exactly the same training set. This can make code errors (bugs) difficult to spot. We verified the network's correct operation by using a baseline training set for each of the network architectures used prior to each experiment. The baseline training set used a unique value for all inputs for each output type, a very simplistic training set. In all cases, we saw the network error descend to a 1% error rate.

Formal neural network training methodology splits the training set into a training section and a testing section. This is because testing on examples you have not trained the network on is a more realistic determination of real world performance.

4. Experiments

4.1 Experiment One

We used two whistle ‘words’ for our initial investigation. The first whistle was simply a peak, a rising then falling pitch. the second consisted of a cycle, where the pitch rose then fell, stayed constant and then finished upward. We extracted 16 examples of a variant of the first whistle (c.f. Figure 4.1). For the second whistle, we extracted 10 examples of one variant (c.f. Figure 4.2) and 9 examples of a second variant (c.f. Figure 4.3).

For testing purposes, we used four of the first whistle examples and two of each variant of the second whistle. In addition, we had a third output class in which the training examples consisted of a variety of non-whistle noise. it is arguable that testing on a whistle variant which is a playback of the same sound upon the network is trained, will

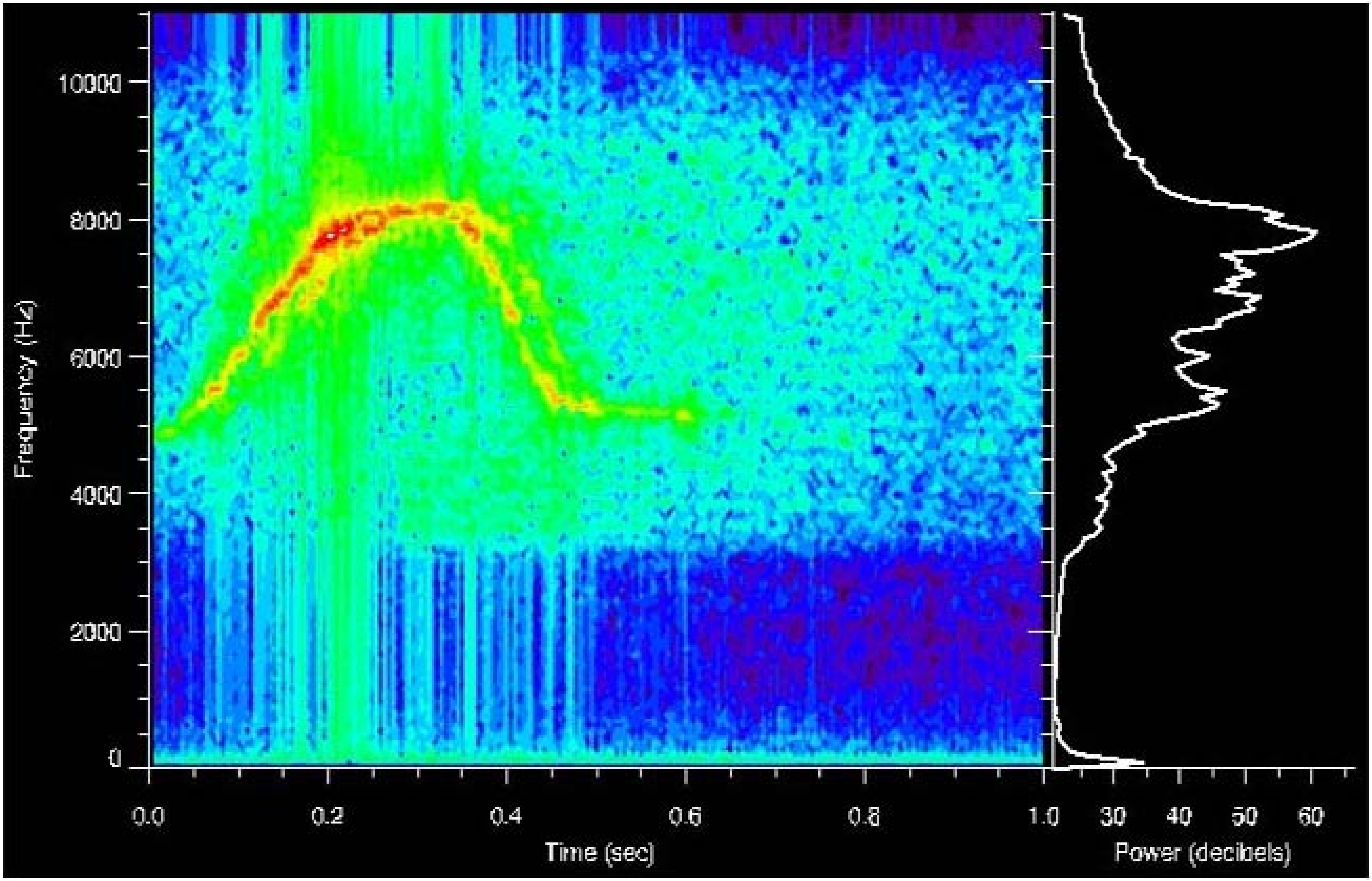

Figure 4.1 - Peak 1

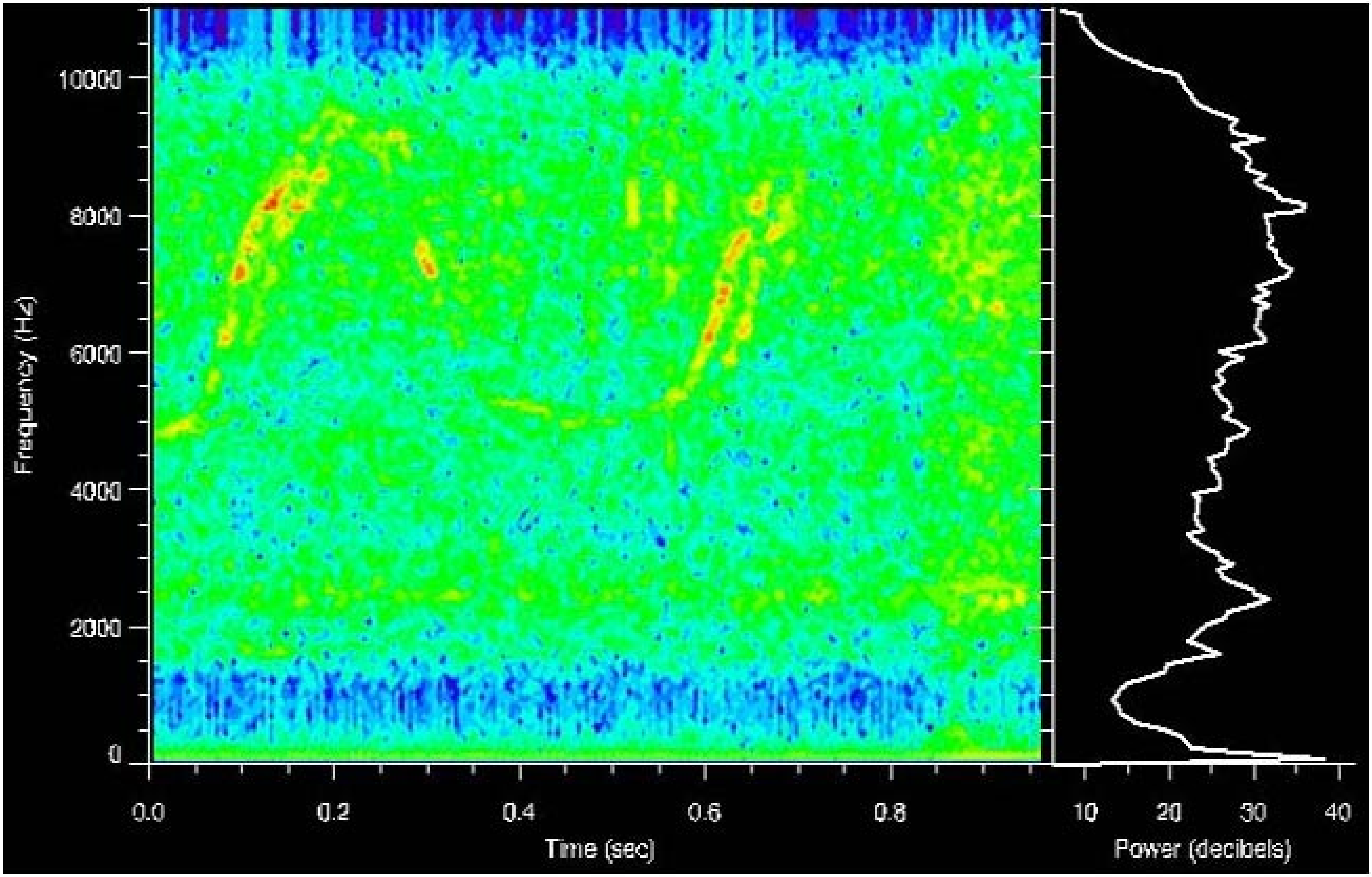

Figure 4.2 - Cycle 1

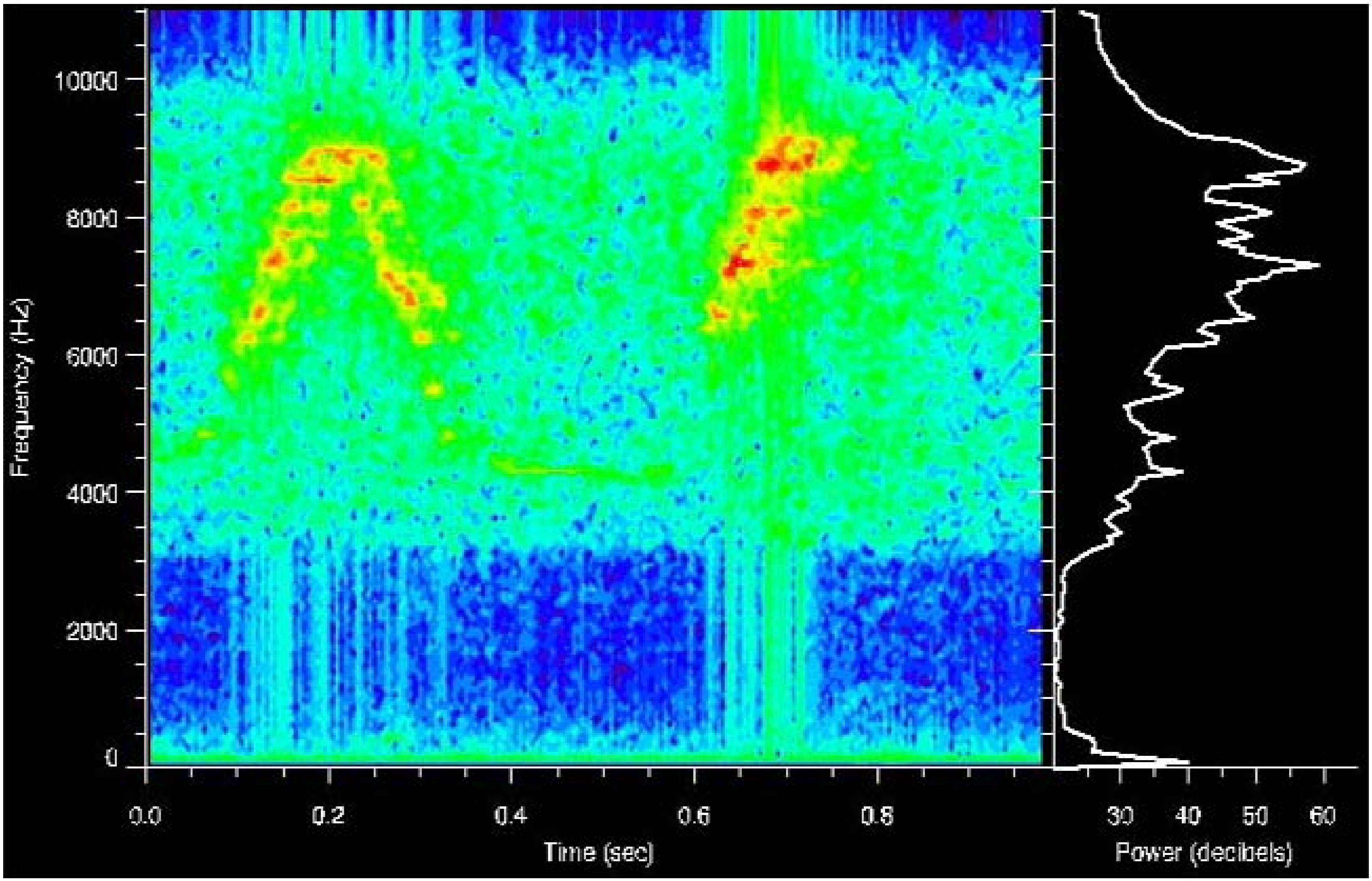

Figure 4.3 - Cycle 2

invariably lead to a strong indicated performance. For this reason, we took all recordings from the dolphin tank into which the sounds were played, not just the sounds directly from the generating computer. This saw additional noise added such as water and dolphin motion noise as well as variation caused by water characteristics.

Additionally, we used two variants of one whistle to determine whether the network could class these variants as one word.

The network architecture used was a three layer network of 16640-20-3 nodes. We saw an initial total sum squared error rate (TSSE) of 27.4% which decreased to 9.9% after 430 training cycles (epochs).

All 12 test examples were correctly classified using the highest node value. The peak outputs of the first whistle (single variant) were in the order of 0.94 to 0.96. The outputs of the cycle whistle (two variants) were in the order of 0.59 to 0.64. Noise output node values ranged from 0.99 to 0.72.

Thus, based on output strength, multiple variants are more difficult to successfully detect (as would be expected) compared to single variants. However for this small initial training set, output node values are strong enough to detect a whistle presence.

4.2 Experiment Two

For our second experiment, we expanded the training set to contain a total of four whistle 'words'. Each word consisted of two variations of a fundamental type. The additional 'peak' variant added is shown in Figure 4.4. The two 'cycle' words used in the first experiment were retained. Two 'up' word variants were added and are shown by Figures 4.5 and 4.6, and two 'down' word variants added and are depicted by Figures 4.7 and 4.8 respectively.

The training set consisted of a total of approximately 20 examples of each word as well as 20 examples of non-words (i.e. other dolphin whistles and environmental noise). Ninety-nine total examples were used for training and 41 for testing.

With this training set, the network demonstrated very poor performance during training (and also in the testing phase). We experimented with a variety of network configurations, modifying the

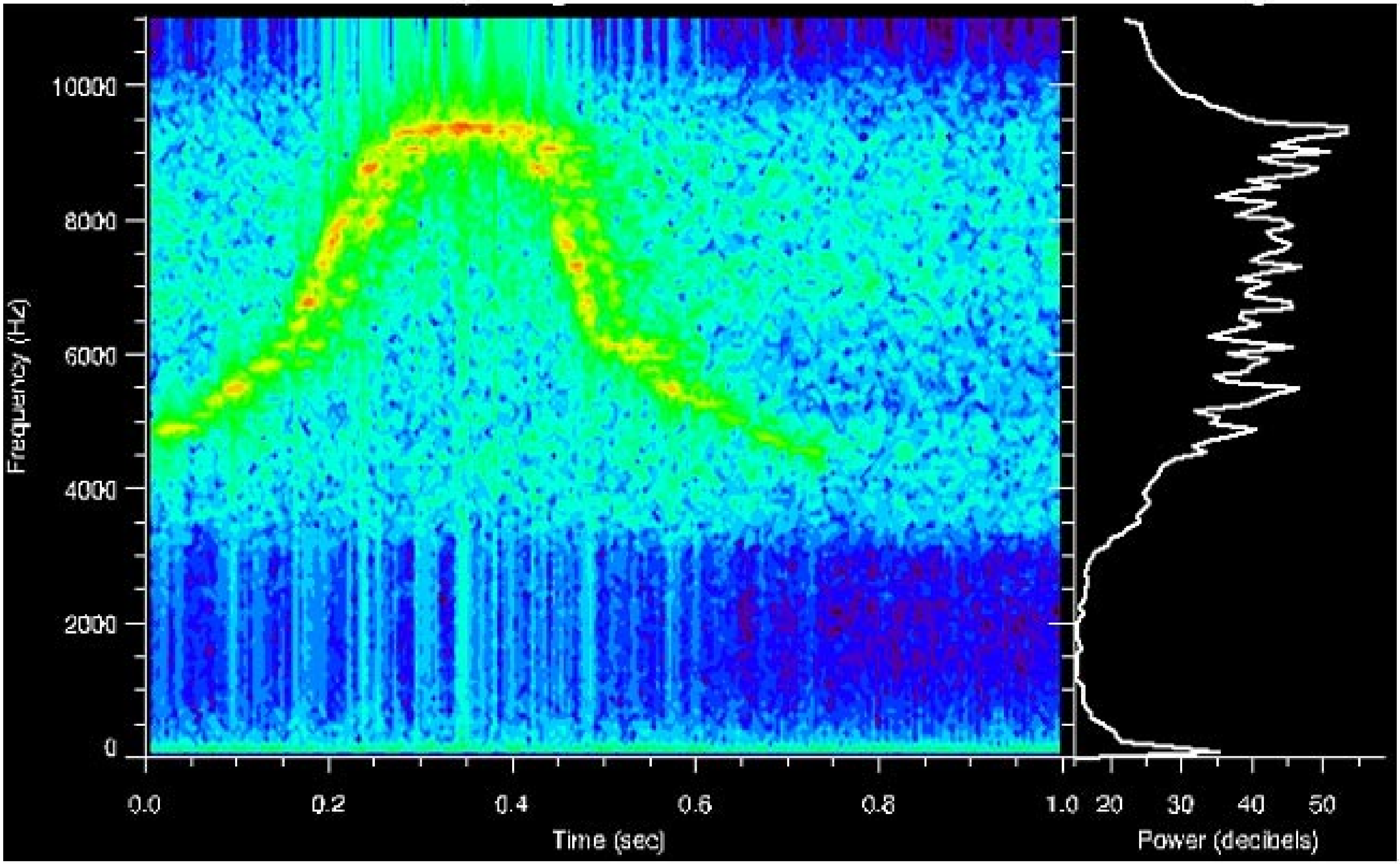

Figure 4.4 - Peak 2

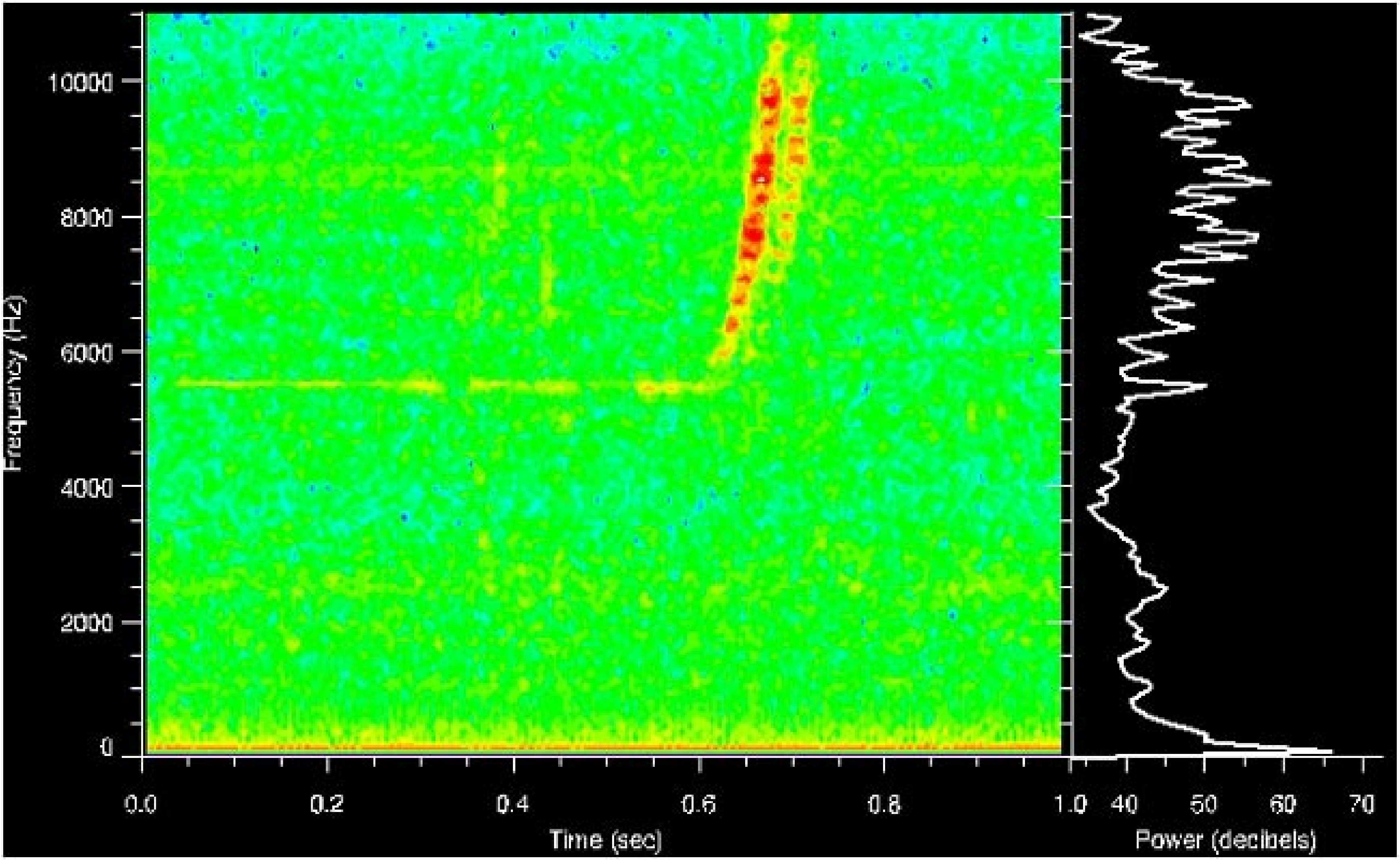

Figure 4.5 - Up 1

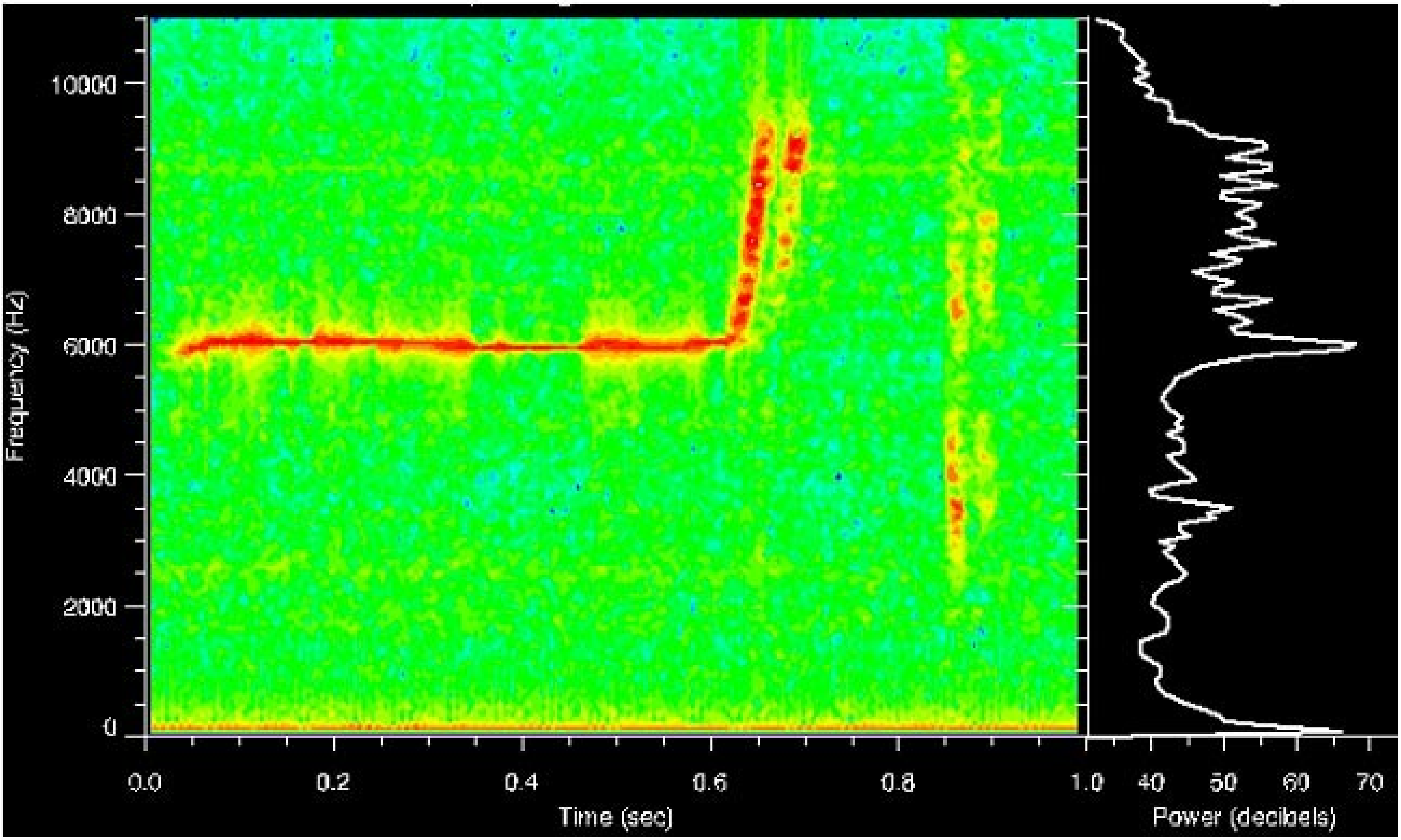

Figure 4.6 - Up 2

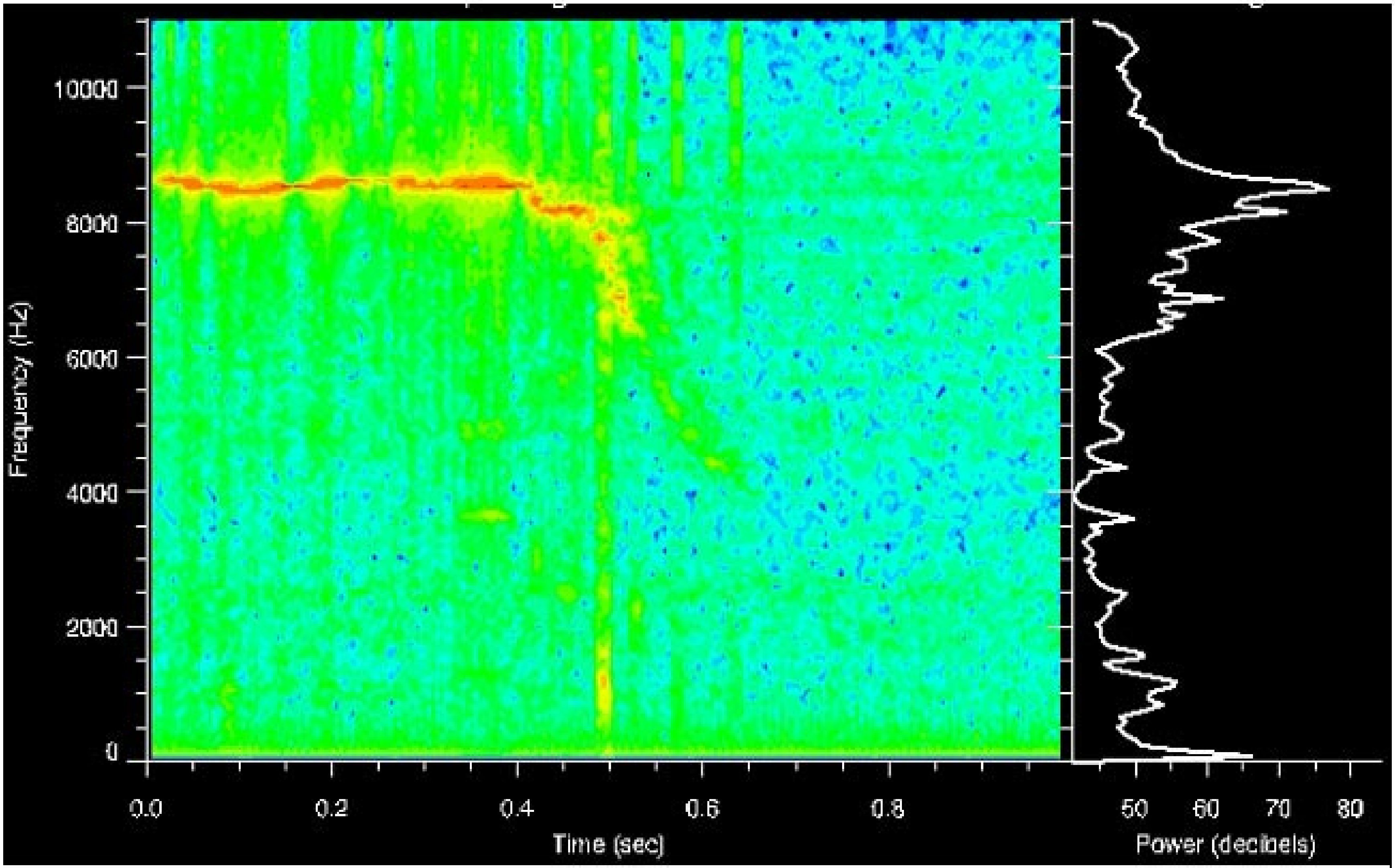

Figure 4.7 - Down 1

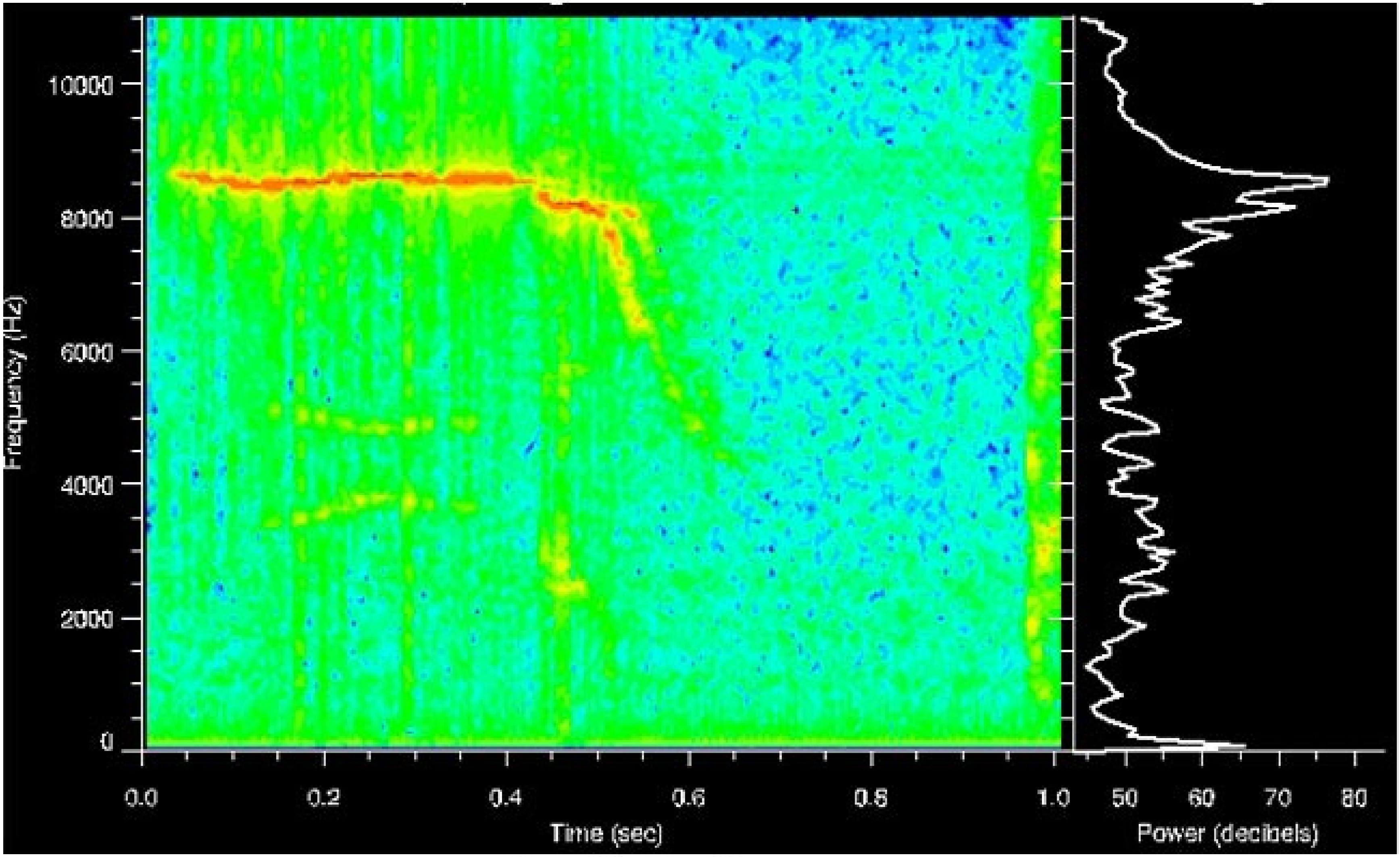

Figure 4.8 - Down 2

number of hidden layer nodes between 15 to 50 nodes. However, apart from an increase in training time, there was no significant improvement in the rate of error. The best result from the network was a training session in which the total sum squared error reduced from 78% to 49% over 500 epochs. The performance on the test set was an error ratio of 19/41 incorrectly classified input patterns.

We then investigated modifying the window sample size for the spectrographic code. We had settled on a sample window of 256 sample values with a step increment of 100 samples, as the resulting 256 frequency bins provided a visual differentiation for the different words. We recreated two additional training sets from the same recordings using a window size of 128 samples and a size of 64 samples. The step increments were halved for each training set in order to retain the same number of input nodes (16640) for the neural network.

The two new training sets both saw a slight drop in the initial TSSE of the first epoch from 79% to 78% and after 500 epochs, the TSSE had dropped to approximately 38% for both training sets. However, test performance remained the same as for the initial training set which was 19 out of 41 patterns incorrectly classified.

For curiosity we also ran the network training epochs to a total of 750 which saw training TSSE decrease an extra 2% to 36% but the same test result occurred.

5. Summary

We saw strong recognition capability with a small training set consisting of two whistles and noise. The network performed well with a single variant but less well with two variants. With an expanded training set of multiple words with multiple variants of each word (in addition to noise), we saw recognition performance decrease dramatically. Variations of the network architecture and variations of the sample window size used for spectrographic analysis made little difference to the end result.

5.1 Conclusions

We conclude that although a backpropagation feed-forward neural network is able to recognize very specific whistle characteristics, the generalization ability is not strong enough to cater for a wider divergence of whistles that are required to be classified as the same category or ‘word’.

5.2 Futures

A constructed whistle recognizer is an essential tool for allowing animal researchers to determine whether a dolphin reproduced whistle is similar to a constructed whistle and if so, what particular whistle is being reproduced. Without such a tool, the only possible clarification method is a visual check of a realtime sonogram which is difficult to accurately do over a lengthy period of time. It also distracts the investigator from the present actualities of the animal experiment in progress.

Thus, a recognizer is an important stepping stone to a consensus language. It would provide a communication path from dolphin to human and overcome the difficulties of translating between two different auditory domains.

Further investigation into neural network use would most probably require expanded pre-processing methods in order to make variant whistle types less divergent so that a neural network could more readily classify the same types.

Appendix A - IDL spectrographic code

SPECTROGRAM.PRO

```
PRO spectrogram, filename

	IF (N_ELEMENTS(filename) LT 1) THEN filename = 'test1.wav'

	wav = READ_WAV(filename, rate)
	dt = 1./rate   ; time spacing
	n = N_ELEMENTS(wav)

	print, rate, n
	WINDOW, XSIZE=800, YSIZE=600
	DEVICE, DECOMPOSED=0
	LOADCT, 39   ; rainbow colors

	!P.FONT = 1    ; truetype font
	!P.CHARSIZE = 1.5   ; larger

; pick some default styles & tick lengths
	!X.STYLE = 9  ; exact X axis
	!Y.STYLE = 9  ; exact Y axis
	!X.TICKLEN = -0.02
	!Y.TICKLEN = -0.02

; choose plot location in normalized coordinates
	x0 = 0.15
	y0 = 0.1
	x1 = 0.75
	y1 = 0.7
	position = [x0, y0, x1, y1]
	alltime = dt*FINDGEN(n)
	PLOT, alltime, wav, $
		POSITION=[x0, y1+0.1, x1, 0.95], $
		XSTYLE=9, YSTYLE=9, $
		XTICKLEN=-0.08, $
```

```
        XTICKNAME=REPLICATE(' ',29), $
        YMINOR=1, $
        TITLE=filename

    nsample = 256  ; could change this to 512
    nfreq = nsample/2 + 1    ; number of frequencies
    skip = 100   ; spacing between samples
    nspec = (n - nsample)/skip
; construct two-dimensional array to hold the results
    spec = FLTARR(nspec, nfreq)
; loop through time
    FOR i=0, nspec-1 DO BEGIN
        spec[i,*] = SPECTRUM(wav[i*skip:i*skip+nsample-1], dt, $
            FREQ=freq, PERIOD=period)
    ENDFOR

; construct time array (the nsample/2 shifts by half the window)
    time = dt*(FINDGEN(nspec)*skip + nsample/2.0)

; Convert power to decibels to make it more visible
    logspec = 10*ALOG10(spec > 0.1)


; Contour plot the spectrogram
    CONTOUR, logspec, time, freq, $
        /FILL, $
        /NOERASE, $
        NLEVELS=20, $
        POSITION=position, $
        XRANGE=[MIN(alltime), MAX(alltime)], $
        XTITLE='Time (sec)', $
        YTITLE='Frequency (Hz)', $
        TITLE='Spectrogram of WAV file'
```

```
; Now plot the spectrogram averaged over time
      global_power = TOTAL(spec, 1)/nspec     ; average over first
dimension
      log_global = 10*ALOG10(global_power > MIN(global_power[1:*]))

      PLOT, log_global, freq, /NOERASE, $
            POSITION=[x1, y0, 0.95, y1], $
            XRANGE=[MIN(log_global), MAX(log_global)*1.1], $
            THICK=2, $
            YTICKNAME=REPLICATE(' ',29), $
            YTICKLEN=-0.08, $
            XMINOR=2, $
            XTITLE='Power (decibels)', $
            TITLE='Average'

; and dump it to a binary file
      OPENW, 1, filename + '.spg'
      WRITEU, 1, nspec
      WRITEU, 1, nfreq
      WRITEU, 1, logspec
      CLOSE, 1


END
```

SPECTRUM.PRO

```
;-------------------------------------------------------------
;+
;   SPECTRUM
;
; PURPOSE:
;   Compute Fourier Power Spectrum
;
;
; CALLING SEQUENCE:
;
;   f = SPECTRUM(x,dt,FREQ=freq,PERIOD=period)
;
;
; INPUTS:
;
;   x = original time series, of length N
;
;   dt = time interval between x measurements = Total time/N
;
;
; OUTPUT:
;
;   F = output power spectrum with N/2+1 components  (units are X^2)
;
;
; OPTIONAL KEYWORD INPUTS:
;
;    FRACTION = output F^2/TOTAL(F^2) (fractional power spectral
density)
;
;   LAG1 = LAG 1 Autocorrelation, used for SIGNIF levels. Default is 0.0
;
;   SIGLVL = significance level to use. Default is 0.95
;
```

```
;   TUKEY = smooth F using a Tukey filter. Default is no smoothing.
;
;   WIDTH = the effective width of the Tukey filter, as a fraction of the
;          number of points N. Default is 0.02
;
;
; OPTIONAL KEYWORD OUTPUTS:
;
;   FREQ = output frequency components [ FREQ(n) = n/(Ndt) ]
;
;   PERIOD = output period components [ PERIOD(n) = 1/FREQ(n) ]
;
;   SIGNIF = output significance levels as a function of FREQ
;
;   AMP = Fourier amplitude
;
;   PHASE = Fourier phase
;
;     BANDWIDTH = The width of the smoothing filter in units of
frequency
;
;      DOF = Degrees of freedom for chi-square distribution for
significances.
;        For no smoothing this is 2.0.
;        For the Tukey it is 2*N*WIDTH
;
;
; MODIFICATION HISTORY:
;   Written C. Torrence
;   5 Aug 1998 (CT): added BOXCAR filter
;   3 Feb 1999 (CT): remove BOXCAR, added TUKEY
;-
;----------------------------------------------------------
```

```
FUNCTION SPECTRUM,x,dt, $
	FRACTION=fraction, $
	LAG1=lag1,SIGLVL=siglvl, $
	TUKEY=tukey,WIDTH=width, $
	FREQ=freq,PERIOD=period, $
	AMP=amp,PHASE=phase, $
	FFT_THEOR=fft_theor,SIGNIF=signif,DOF=dof, $
	BANDWIDTH=bandwidth

	ON_ERROR,2
	IF (N_ELEMENTS(siglvl) LT 1) THEN siglvl = 0.95
	IF (N_ELEMENTS(lag1) LT 1) THEN lag1 = 0.0
	IF (N_ELEMENTS(width) LT 1) THEN width = 0.02

	N = N_ELEMENTS(x)
	fft_x = FFT( x(*) - TOTAL(x)/N ,-1)
	amp = 2*ABS(fft_x[0:(N+1)/2-1])
	IF ((N MOD 2) EQ 0) THEN $  ;Nyquist is 1/2 power for N even
		amp = [amp,SQRT(2)*ABS(fft_x[N/2])]
	phase = (ATAN(IMAGINARY(fft_x),FLOAT(fft_x)))[0:(N+1)/2-1]
	power_spec = 0.5*amp^2
	variance = TOTAL(power_spec[1:*])

	IF KEYWORD_SET(fraction) THEN fraction=1./variance ELSE
fraction=1

	nf = N/2 + 1
	freq=FINDGEN(nf)/(N*dt)
	period=[N*dt,1./freq(1:*)]

	dof = 2.
	IF KEYWORD_SET(tukey) THEN BEGIN
		m = (3./4)*width*N
		power_spec =
FILTER_TUKEY(power_spec,m,N*dt,bandwidth,dof)
	ENDIF
```

```
	fft_theor = (1 - lag1^2)/(1 - 2*lag1*COS(dt*freq*2*!PI) + lag1^2)
	fft_theor = fft_theor*(variance*2./N*fraction)
	signif = fft_theor*(CHISQR_CVF(1.-siglvl,FLOAT(dof))/FLOAT(dof))

	RETURN,power_spec*fraction
END
```

Bibliography

[1] Albers, Vernon M
The World of Sound 1971

[2] Au, W. W. L, Andersen L. N, Rasmussen A. R, Roitblat H. L, Nachtigall P.E
Neural network modelling of a dolphin's sonar discrimination capabilities
J. Acoust. Soc. 1998

[3] Buck J.R, Tyack P.L
A quantitative measure of similarity for tursiops truncatus signature whistles
J. Acoust. Soc. Am. Vol 94 No 5 Nov 1993

[4] Caldwell M, Caldwell D - Vocal mimicry in the whistle made by an atlantic bottlenosed dolphin
Cetology 1972

[5] Chang W, Bosworth B and Clifford Carter G -
Results of using an artificial neural network to distinguish single echoes
from multiple sonar echoes
Naval Undersea Warfare Center, Connecticut
J. Acoust. Soc. Am. Vol 94 No 3 Pt 1 Sep 1993

[6] Garrison, Tom - Oceanography
Wadsworth Publishing, 3rd Edition, 1999

[7] Herman, Lou - Cetacean Behavior: Mechanisms and Functions
University of Hawaii, Honolulu, Hawaii 1980
John Wiley & Sons, New York

[8] Herman L, Richards D, Wolz J
Comprehension of sentences by bottlenosed dolphins
Cognition, Vol 16 p129-219 1984

[9] Herman L.M, Kuczaj S.A, Holder M.D
Responses to Anomalous Gestures by a language trained dolphin - evidence for processing of Semantic relations and syntactic information
J. Experimental Psych. General Vol 122 Iss 2 pp184-194 1993

[10] Houser et al - Echolocation click classification
J. Acoust. Soc. Am Vol 106, Sep 1999

[11] Houser D. S, Helweg D. A, Moore P.W.
Classification of dolphin echolocation clicks by energy and frequency distributions
J. Acoust. Soc. Am. Vol 106 No. 3 Pt 1 Sep 1999

[12] Janik V.M, Slater P.J
Context-specific use suggests that bottlenosed dolphin signature whistles are cohesion calls
Animal Behaviour, Vol 56 pp 829-838 1998

[13] Janik V.M.
Pitfalls in the categorization of behaviour -
a comparison of dolphin whistle classification methods
Animal Behaviour Vol 57 p133-143 1998

[14] Kadtke J, Kremliovsky M
Estimating dynamical models using generalized correlation functions
Marine Physical Laboratory, UC San Diego CA
Physics Letters A 260 p203-208 1999

[15] Meister J
A Neural Network harmonic family classifier
Diagnostic/Retrieval Systems Inc New Jersey
J. Acoust. Soc. Am. Vol 93 No 3 Mar 1993

[16] Murray S.O, Mercado E, Roitblat H.L
Characterizing the graded structure of false killer whale vocalizations
J. Acoust. Soc. Am. Vol 104 Pt 1 Sep 1998

[17] Pepperberg, I.M.: A communicative approach to animal cognition: A study of conceptual abilities of an African Grey parrot (Psittacus erithacus).
Cognitive Ethology: The Minds of Other Animals.
Hillsdale, NJ, Erlbaum Assoc., 1991.

[18] Rahim M. G, Goodyear C. C, Kleijn W. B, Schroeter J, Sondhi M. M.
On the use of neural networks in articulatory speech synthesis
J. Acoust. Soc. Am. Vol 93 No.2 Feb 1993

[19] Ralston J, Herman L. M.
Perception and generalization of frequency contours by a bottlenose dolphin
J. of Comparative Psych. Vol 109 pp268-277

[20] Reiss D, McCowan B
Spontaneous vocal mimicry and production by bottlenosed dolphins - evidence for vocal learning
J. Comparative Psych. Vol 107 Iss 3 pp 301-312 1993

[21] Roitblat H. L, Moore P. W. B, Helweg D. A, Nachtigall P. E
Representation and processing of acoustic information in a biomimetic network
Animals to Animals 2
Proceedings of the Second International Conference on Simulation of Adaptive Behaviour, MIT Press, Cambridge pp 90-99 1992

[22] Savage-Rumbaugh S, Shanker S. G, Taylor T. J
Apes, Language and the Human Mind
Oxford University Press 1998

[23] Sayigh L.S, Tyack P. L, Wells R. S, Solows A. R, Scott M. D, Irvine A. B.
Individual recognition in wild bottlenose dolphins:
a field test using playback experiments
Animal Behavior 57 p41-50 1999

[24] Sigurdson J.
Frequency-modulated whistles as a medium for communication with the bottlenose dolphin
Language and Communication: Comparative Perspectives
pp 153-173 Hillsdale, New Jersey 1993

[25] Svardstom A
Neural network feature vectors for sonar targets classifications

J. Acoust. Soc. Am Vol 93 No 5 May 1993

[26] Workman G. L.
An acoustic emission and acousto-ultrasonic analysis of impact damaged composite pressure vessels
NASA Technical Report CR-203493 Jan 1996