Extraction and Recognition of Tonal Sounds Produced By Small Cetaceans and Identification of Individuals

by

C.R. Sturtivant

Doctoral Thesis

Submitted in partial fulfilment of the requirements

for the award of

Doctor of Philosophy of Loughborough University

November 1997

Table of Contents

Synopsis

The by-catch of small cetaceans in fishing nets has been identified as a widespread problem, but attempts to reduce this require an understanding of the way these animals behave around the nets. One of the problems with assessing changes in behaviour between encounters is the difficulty of identifying individuals. Acoustic identification techniques overcome some of the problems associated with visual ID, and field research has shown that the presence of a sonobuoy and hydrophone have no effect on dolphin behaviour in the field. Dolphins produce whistles that can be used for identification, although current theory suggests these identify small groups rather than individuals.

Novel algorithms have been developed to detect and process these tonal whistles, and their characteristic time-frequency-intensity contours extracted from the raw signals. Feature extraction techniques were developed for the contours based on time-frequency 'shape' of the contours, allowing a syntactic pattern recognition approach based around hidden Markov modelling to be employed for classification.

The algorithms have enabled the whistles from concurrent whistles to be separated and analysed. Contours of 101 wild bottlenose dolphin whistles were successfully characterised. Analysis of the resulting classes indicated one group occurring only once and two other groups occurred twice but on different days. Another study was conducted of three groups of common dolphin, with a total of 49 recorded whistles analysed. The first group was found to contain whistles significantly different to either of the other two, although neither similarity nor dissimilarity could be inferred on the second and third. Further analysis suggested there were indeed two separate groups of dolphins for the last two groups, but that there was a period of overlap in their recording. A significant difference could be found between them once certain classes were re-assigned.

It should be possible to apply these same techniques to a wider range of odontocete species, since most of those studied have been found to exhibit similar whistles. The tasks of whistle detection, isolation, and encoding can be applied automatically by computer with no loss of identity information, and these encoded contours can subsequently be quantitatively classified by their shape.

Chapter 1: Introduction

Background

From 1991 to 1993 Loughborough and Cambridge universities conducted field trials of an acoustic reflector in the Moray Firth, Scotland. The reflector was a prototype of a design that could be fitted to fishing nets and would reflect back a dolphin's echolocation pulse, thereby alerting it to the presence of an obstruction in the water. The commonly used mono-filament nets are made of single strands of nylon, a material with similar acoustic properties to that of water, and which thus returned an extremely weak echo back to the dolphin.

The Moray Firth is inhabited by a group of bottlenose dolphins (*Tursiops truncatus*), which although they are not one of the species at greatest risk from net entanglement, are of approximately the same size as the common dolphin (*Delphinus delphis*) (Figure 1-1) and employ a similar echolocation signal. It is the common dolphin and the harbour porpoise (*Phocoena phocoena*) which are the two species most frequently found entangled in nets around the UK coast, and there are local European populations which are in danger of extinction from this problem. Harbour porpoises also frequent the Moray Firth and Cromarty Firth, but are much smaller than the bottlenose and the common dolphins. The length of a harbour porpoise is 1.4–1.7 metres, whereas that of the common dolphin is 1.7–2.4 metres, and the bottlenose dolphin is 2.9–4.0 metres. Most of the adult bottlenose dolphins in the Moray Firth are close to 4.0 metres, since this latitude is near to their maximum northern range. (Distribution and sizes from Martin, 1990.)

A position about a mile south of Cromarty was chosen for the study, as it afforded a good vantage-point from a 50 metre high cliff and was accessible with four-wheel drive vehicles. Reports from a research team based at Aberdeen University, who were conducting a study on the local population using photo-id, suggested that the area was commonly traversed by the dolphins. More than 100 bottlenose dolphins have been identified in the study area (Wilson *et al.*, 1993), and the total population size is estimated to be 142 individuals (Wilson, *pers. comm.*). The cliff top observations suggested that dolphins generally travelled through the area,
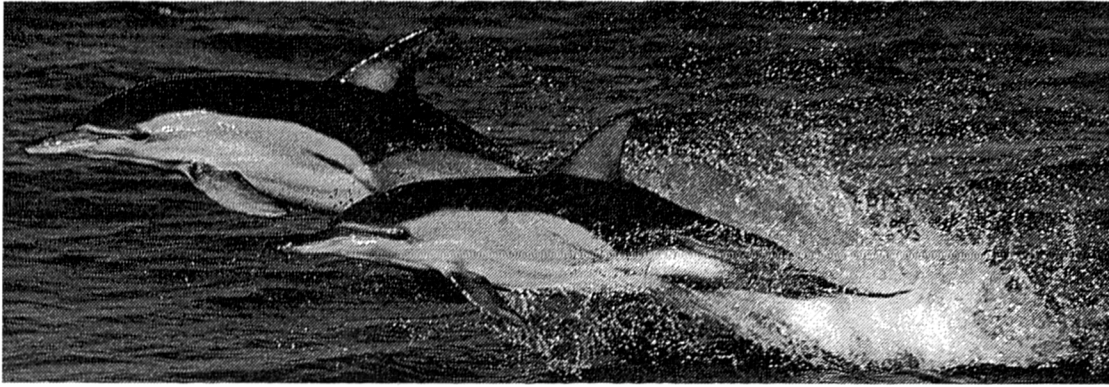
Figure 1-1: Common Dolphins (*Delphinus delphis*)

rather than staying to forage, although foraging behaviour was observed on a small number of occasions.

The monitoring equipment and test devices were deployed away from the shore, so that breaking waves would not cause air to be entrained in the water and therefore avoiding the possibility of damage. The line of acoustic reflectors was arranged in a grid pattern with varied spacing on different trials, and was placed so as to intercept the dolphins' favoured route. Individuals were tracked using a theodolite equipped with an electronic data logger, and their vocalisations monitored with hydrophones connected either to sonobuoys or directly cabled to recording equipment.

The purpose of the recordings was to monitor the dolphins' echolocation activity, but soon after the start of the first trial it also became useful as an early warning system. Frequently a group of whistling dolphins could be heard on monitoring headphones long before they came into sight. After the trials had finished it was suggested that it would be useful to know the identity of each group of dolphins as it travelled through the area, and it was these whistles that seemed to hold the key to identifying them.

The Study of Dolphin Bio-Acoustics

From the first occasion when small cetaceans were kept successfully for extended periods of time at Marine Studios in Florida in the 1930's, dolphins were discovered to be very vocal animals. The bottlenose dolphins that were maintained there (often confusingly referred to in early literature as 'the porpoise') were initially kept merely as just another example of local marine life. However, their apparent curiosity and

3

willingness to interact with people led to them becoming firm favourites with visitors and staff alike. The facial anatomy of this species, which conveys the appearance of a permanent (and often misleading) smile, almost certainly contributed to this popularity, and soon the dolphin had caught the attention of researchers and the public who both wanted to learn more about them.

In a discussion of the history of Marine Studios (which later became Marineland of Florida), Forest Wood wrote:

> "The opportunity to observe these exotic animals at close hand dispelled some erroneous beliefs … and provided a host of new information. It was apparent that the porpoises had good vision in air as well as water; that they made a variety of sounds, especially when excited; they slept in brief naps at intervals during the day and night; and if there was a current in the tank they continued to swim while sleeping; they exhibited a wide and imaginative variety of sexual behaviors." (Wood, 1973)

The earliest attempts at studying the dolphins were mainly in the area of husbandry and vocal behaviour. Wood attempted to describe some of the noises they produced, using words such as 'whistles, rasping and grating sounds, combined mewing and rasping sounds, barks, and yelps', which gives some indication of the range and types of vocalisations. Clearly these animals had a large vocal repertoire, suggesting that sound might form an important part of communication within their social systems. Also, the dolphins were often curious about man-made noises and would on occasion mimic sounds made by people around their pool.

Echolocation Research

It was at Marine Studios that observers first suspected that dolphins could use sound to gain information about their surroundings. In 1956 Arthur McBride, who was then curator at Marine Studios, speculated on the use of sound by wild dolphins when he was attempting to catch some for display and they consistently managed to avoid the nets he had set for them:

> "May we not suspect that the above described behavior is associated with some highly specialized mechanism enabling the porpoise to learn a great deal about his environment through sound?" (McBride, 1956)

4

Wood, who took on the job of curator after McBride, noticed that whenever some strange object was introduced into the pool, the dolphins would produce the same sorts of 'rasping sounds' when they came over to investigate it:

"The general pattern of behavior under these circumstances is always the same; the porpoises bunch together and swim at a faster rate around the tank. They accelerate while passing the object, but crane their necks to look at it as they go by. Rasping sounds, usually brief, are heard at this time. Generally within only a few minutes (the Marineland porpoises have been subjected to many unfamiliar objects) their swimming slows and the younger animals (those which were born in the tank) approach the object more closely, and even pause and appear to examine it. If it is the hydrophone they are examining, the "rusty hinge" noise and related rasping or grating sqund are, of course, quite pronounced when heard through the listening equipment." (Wood, 1953).

As well as commenting on the occurrence of these sounds, Wood made suggestions about the purpose to which these sounds might be put:

"[These sounds] possess the characteristics necessary for a pulse-modulated type of echo-ranging, and the behavior of Tursiops at the time it emits these sounds certainly suggests that they might be used for echolocation or even what might be termed 'echo-investigation'." (Wood, 1953)

Schevill & Lawrence (1953) put forward a similar theory for echolocation in dolphins as an aid for food-finding, as well as Kellogg (Kellogg *et al.*, 1953; Kellogg 1958) who went on to publish one of the first books on dolphin echolocation 'Porpoises and Sonar' (Kellogg, 1961). By comparison, the first paper published on the sonar of bats was by D.R. Griffin in 1938, some fifteen years earlier, but the first textbook on bat echolocation (Griffin, 1958) did not appear until just three years before Kellogg's one on dolphins.

Dolphin Whistles Research

In his 1953 paper, Wood also suggested a potential use for the whistles he had heard during his observations:

"There appears to be good evidence that whistling may constitute a form of communication between a mother porpoise and her infant, enabling one to find the other in the event of separation." (Wood, 1953)
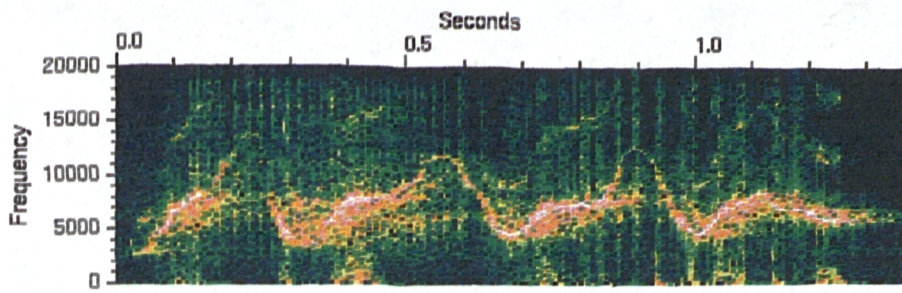
Figure 1-1: Typical spectrogram of a bottlenose dolphin whistle and echolocation clicks

First descriptions of some of the characteristics of the bottlenose dolphins' whistles (Kellogg *et al.*, 1953) reported their frequency range as between 7 and 15 kHz, and also noted that whistles tended to vary considerably between different individuals. The frequency range of the whistles has been variously stated in subsequent papers as 4–15 kHz (Dreher, 1961), 4-18 kHz (Lilly & Miller, 1961), 4–20 kHz (Evans & Prescott, 1962), 5 16 kHz (Dreher & Evans, 1964), and 5–15 kHz (Herman & Tavolga, 1980) (from Caldwell *et al.*, 1990). It is possible that the range may vary with geographical location, and whistles from this species have been recorded containing frequencies up to 30 kHz (Ridgway *pers. comm.*, 1996). Geographical difference might explain why the quoted frequency range varies so much, although equipment limitations may also have had some effect.

Kellogg's observation in his 1953 paper that whistles appeared to differ between individual dolphins was taken further in 1965 by David and Melba Caldwell (Caldwell & Caldwell, 1965). They put forward the hypothesis that dolphins used certain whistles that distinguished them from other dolphins, and they termed these 'signature whistles'. It seems that the hypothesis itself is sometimes misrepresented, so it is worth including their explanation from their review of the signature-whistle hypothesis for the bottlenose dolphin:

> "This hypothesis arose from our observation, made in 1965, that each dolphin in a captive group tended to produce whistles which were individually distinctive and stereotyped in certain acoustic features. Since the individually distinctive features of these whistles were so striking to us, we called them signature whistles. While some acoustic features of an individual's whistles change as a function of behavioral context, as discussed below, others do not. We have hypothesized that the individually distinctive attributes of signature whistles function to broadcast the identity of the whistler. The more variable acoustic features of the

The use of spectrograms (Figure 1-2) to plot the time-frequency contour of a whistle resulted in an increase in the study in this area, but there was no easy way available at the time to do all but simple comparisons of the whistles. For this reason much of the early results were based on subjective, visual comparisons of whistle spectrograms. Later attempts to classify the whistle by examining the number of 'loops' (where a whistle's spectrogram showed several 'U'-shaped elements closely spaced in time) were more objective, but still lacked inclusion of any of the details of a whistle's time-frequency contour in the comparison (see Caldwell *et al.*, 1990).

Furthermore, it seems possible that these 'loops' used in the analysis formed part of continuous whistles, the problem lying in the fact that various parts of the apparatus might have been insensitive to the higher frequencies contained in the whistle. Watkins (1974) has reported such problems when recording cetacean vocalisations, although in his paper he concentrated on recordings of echolocation signals. The result of insufficient sensitivity at the upper end of the sound spectrum would be a spectrogram in which the whistle was shown with the high frequency 'peaks' missing, thus giving the impression of a series of shorter whistles in close proximity to each other. Early printed spectrograms did little to help, since the range of ink shadings for different sound intensities was very small. Caldwell *et al.* refer briefly to this problem:

"For some dolphins there is often no detectable sound between loops; however, the intervals between loops are relatively constant. ... For example, Schevill and Watkins (1962, their Fig. 13) illustrated what we would call a single multi-looped whistle, even though the spectrogram indicates silent periods between repeats. In the Tavolga (1968) figure of what may well be the same animal ..., a signal is detectable between cycles on the sound spectrogram, but it is faint. Lang and Smith (1965) discussed this problem in presenting their results and resolved it by dotting in lines as they perceived them by ear." (Caldwell *et al.*, 1990)

This type of 'interrupted' whistle has however been observed in our recordings of the wild dolphin Freddie from Amble, Northumberland on which a number of studies have been conducted (Bloom, 1991; Bloom, 1992; Bloom *et al.*, 1995). Caldwell *et al.* indicate that their data has been collected starting around 1965, which

7

suggests high frequency recording problems may have occurred. In their Results section (Caldwell *et al.*, 1990 Figure 1), the spectrograms for all their whistles show a marked decrease in intensity starting at approximately 15 kHz, and no whistles are shown where the upper loop is present. This factor suggests that the 'loop' explanation could be due to the recording procedures rather than any real phenomenon. In the same paper (Caldwell *et al.*, 1990 Figure 17), a lower frequency whistle is shown with 'about 2.5 loops' and a maximum frequency of 10 kHz, but where the contour is unbroken.

Later analysis of whistles included such additional parameters as lowest and highest frequencies, whistle duration, and total frequency modulation, but still relied on human assessment to measure these parameters. The following of a time-frequency contour on a spectrogram is not especially difficult when recordings are made in the quiet environments of an oceanarium. When one is presented with a recording made in less than optimal conditions at sea, however, then this task is made much more difficult, and human subjective decisions are required. Difficulties with extraneous noise can clutter the spectrogram to such an extent that it is not easy to tell when one whistle ends and another begins. Such noises include multiple vocalisations from different animals, wave noise, pulsed noises (often referred to as 'squeaks', 'squawks', 'creaks', etc.), boat noise, sounds from other marine species, in addition to other noise sources.

Although the frequency modulations made by the dolphin during a whistle are the most obvious difference between one dolphin and another, it is possible that information important to the dolphin is contained within other attributes. Clicks are sometimes reliably emitted synchronously with a whistle by the same dolphin. This is possible since their nasal passages diverge before the airways reach the nasal sacs, which is from where most researchers believe that these vocalisations are produced. Variations in overall volume in addition to amplitude modulation of the whistle signal have also been observed.

After the formation of the initial signature whistle hypothesis, the literature is rather sparse on the details of whistle structure. This may be due to the difficulty of analysis beyond that already conducted by the Caldwells. There was, however, some research carried out at the Kewalo Basin Marine Mammal Lab, Hawaii, into language

competencies and cognition that among other topics studied the way that bottlenose dolphins perceive differences in whistles (Herman, 1975; Herman *et al.*, 1984; Richards *et al.*, 1984). The most extensive recent study of signature whistles has been carried out by Peter Tyack, who has designed a 'Vocalight' device which can be attached to a dolphin's melon to allow identification of vocalising individuals (Tyack, 1985). He later used this device to show that bottlenose dolphins sometimes mimic each other's signature whistles when interacting (Tyack, 1986; Tyack, 1989).

One might have expected that the more widespread use of affordable computers would have had an impact on the analysis of dolphin vocalisations. In fact it is only recently that they have been used to any large extent in this field (e.g. McCowan, 1995; McCowan & Reiss, 1995), and then only as a measuring tool. Analysis of differences in the detailed structure of whistle contours necessitates objectivity and speed of calculation, which is now provided by computers, and it is possibly for this reason that this area has not been more thoroughly researched by traditional behavioural scientists.

Proposed Application of the Signature Whistle Hypothesis

The Importance of Group Identification

During analysis of the data from the Moray Firth field trials, subjective assessment appeared to suggest that as the trials continued there was an increase in the distance from the acoustic reflector grid at which a group of dolphins would take avoiding action. If we consider the use to which the bottlenose dolphin puts its echolocation capability, then we can form a hypothesis explaining this increase. Acoustic monitoring of the solitary dolphin 'Freddie' revealed that when the animal was not foraging he generally did not use his sonar (Bloom *et al.*, 1995). This animal had a very small, triangular home range of approximately 0.5 km², and was acoustically monitored over several 24 hour periods. Resting and travelling over known territory were not generally accompanied by echolocation activity, however when foraging at a river mouth the dolphin always used his sonar.

The energy cost to the dolphin of echolocating was recently calculated at 83% of the basal metabolic rate (Kole & Speakman, 1993), which is already high due to

this mammal's cold aquatic environment. (Kole and Speakman calculated a basal metabolic rate 75% above that expected for a land mammal of the same size.) Hence, we would expect dolphins to use echolocation only when the need justified this rather high energy cost. This theory was supported from monitoring the vocalisations made by the wild dolphins early in each of the Moray Firth trials, since dolphins in a travel pattern seemed to emit fewer echolocation clicks when they entered the trial area than those that appeared to be foraging. This observation, however, was made rather subjectively, since the amount of echolocation activity heard from the hydrophones had an influence on how the dolphins' activity was classed.

The inter-click interval for Freddie, at Amble, was used to calculate the maximum search range for two-way travel time (see Penner, 1988; Goodson & Datta, 1992), and indicated he was searching for fish out to a distance of no more than 70m. Calculation of the target strength of the size of fish that he was observed to catch, and comparison of this with the source level of his echolocation clicks, suggested that this distance was Freddie's maximum detection range for prey of a normal size. The acoustic reflectors used in the Moray Firth trials were designed to have similar target strengths to that of a large prey fish, and the Moray Firth dolphins initially began taking avoiding action at distances from upwards of 50 metres. The average distance at which avoidance first occurred increased as the trials progressed. This effect might be due to the dolphins increasing the time interval between clicks in the area of the barriers, and thus allowing them to receive echoes from a longer range before the next click overwhelmed further returns.

In the early part of a trial when a group of dolphins would approach the reflector grid, the observers would often see the dolphins pause at a distance and slap their tail flukes on the water surface whilst echolocating continuously. The tail-slapping appeared to occur more often when calves were present in a group, and would frequently be accompanied by the group drawing itself more tightly together. In writing on this subject, Defran and Pryor (Defran & Pryor, 1980) class tail-slaps as an indicator of 'Fear, Stress, or Subordination', and Herman and Tavolga state:

"Most observers agree that tail-slaps convey threat or accompany frustration (e.g., D. Caldwell & Caldwell, 1977), in addition to establishing contact." (Herman & Tavolga, 1980)

The tail-slaps that were observed through the theodolite had the appearance of being rather tentative, and so their interpretation as a sign of frustration or nervousness seems to fit. During later passages through the area, the dolphins would change their direction of travel at longer ranges, and would come to a halt on increasingly rare occasions. An interpretation was developed that those interactions at the commencement of the trails could be described as 'naive' interactions. Those later on were indicative of the dolphins recalling previous, stressful encounters in the area, and that they subsequently were using an echolocation mode of a longer range than they would otherwise have used for simple foraging activity.

The acoustic reflectors would be attached to fishing nets that might stay in the sea for several days at a time, so it was important to gather information on how the dolphins' long term behaviour changed as they became habituated to their presence in an area. In order for this to be possible, the reappearance in the trial area of any group had to be identified.

Photo-ID versus Signature Whistles

Analysis of the recordings indicated the expected decreased pulse repetition rate when avoidance was being made at longer ranges. However, there was no obvious method to prove that the dolphins were indeed increasing their maximum detection range due to previously encountering the barrier at this location. The identity of each group of dolphins remained unknown. Photo-id techniques seemed impractical, since the dolphins were generally approximately 400 to 500 metres from the cliff-top observation positions. In addition, during travel the dolphins would typically remain submerged for approximately 30 seconds and would then only appear at the surface for between half and one second to take a breath before again being lost from view. Even with a powerful camera lens, in this case the chance is small of taking a clear photograph of a dorsal fin for identification, even with an experienced observer who has some idea of where the dolphins are likely next to appear.

Analysis of the dolphins' whistles seemed to be a possible method of determining group identities. The large coverage of the monitoring and recording equipment was found to be excellent for the detection of whistles before any dolphins had been sighted. On one flat calm day, dolphins were reported by boat radio to be

present between the Sutors at the entrance to Cromarty Firth, a distance away of approximately two kilometres, and whistles from these dolphins could be heard over the hydrophones. All visual and acoustic observations were logged with times and dates, so identification of a signature whistle from two different times would suggest that the same dolphin was present on both occasions. Studies using photo-id of changes in group composition by researchers from Aberdeen University (Wilson *et al.*, 1993) showed only a slow change over time. Since each trial lasted a maximum of two weeks, the assumption was made that there would be little change in group composition during each trial period. However, no such assumption about trials from different years could be made.

Whistle mimicry was suggested as a possible drawback to this technique, but when other researchers (Tyack, 1986; Caldwell & Caldwell, 1972; Reiss & McCowan, 1993) have reported this, the mimicry has always taken place within the same social group. Thus, although use of whistle contours alone may not be enough to identify the individual vocalising animal, it can still be used to determine the group to which that animal belongs.

If resources had been available at the start of the trials in 1991, it might have been possible for an observer sitting in a kayak to photograph the dolphins' dorsal fins and skin markings for later identification. The obvious disadvantage to this is that the behaviour of the dolphins might be altered by the presence of the photographer and his boat, and hence the trial might record the reaction of the dolphins to the observers more than to the reflector barriers. Additionally, the tidal flow in the Moray Firth was quite fast, as it is in many coastal locations, and this would have necessitated either mooring an observation boat or paddling out into the trial area whenever a group of dolphins was first sighted.

In conclusion, signature whistle analysis appeared to be the only feasible method of identifying the dolphin groups from these trials, since no sufficiently high resolution photographic data had been collected. If further trials were to be conducted, it is again unlikely that photo-id would be a suitable technique for identification due to safety issues, and possible contamination of the trial results. Little research has been published in the literature on quantitative techniques for the

comparison of signature whistles, and so the development of a research tool to achieve this aim was necessary.

Determining the Identity of Dolphin Groups

Applicability of Signature Whistle Analysis to Other Odontocetes

Bottlenose dolphins are not the only cetacean species that appear to use signature whistles as a means of communication. Many odontocetes exhibit similar stereotypical FM whistles that may be unique to the vocalising individual, although only relatively few of these have been sufficiently well studied for signature whistles to have been proved. Due to their availability for close observation in oceanariums, both the Pacific white-sided dolphin (*Lagenorhynchus obliquidens*) (Caldwell & Caldwell, 1971), and the spotted dolphin (*Stenella plagiodon*) (Caldwell *et al.*, 1973) have been reported to use signature whistles. In addition, the killer whale (*Orcinus orca*) uses pod-specific vocalisations, which have been used as a means to identify pods by researchers on an intensively studied population off the coast of British Columbia (Ford & Fisher, 1981; Dahlheim, 1981; Ford & Fisher, 1983). If a fast, automated, quantitative measure of the similarity between whistles of bottlenose dolphins can be developed, then the same techniques could also be applied to these other species. Since there have been opportunities for detailed study of only a few species, it seems likely that many more odontocetes species might make use of signature whistles or similar vocalisations. If this is the case then these identification techniques could be generally applied in the field.

The most commonly encountered species during the acoustic reflector trials in the Moray Firth was the bottlenose dolphin which, as previously stated, has been reported to use whistles specific to individuals. Among the species believed to be at greatest risk to entanglement in fishing gear is the common dolphin, which although it has been recorded as producing a plethora of FM whistles, no papers have been found in published literature to indicate if these vocalisations constitute signature whistles. It would, however, be surprising if this species did not exhibit similar signature whistle behaviour to that of the bottlenose, spotted, and white-sided dolphins, due to its

similar gregarious and highly vocal nature. It may only have been lack of opportunity that has prevented its use of signature whistles from being discovered.

A second species at risk from bycatch is the harbour porpoise (*Phocoena phocoena*). Although it is still a highly vocal animal when in human care and issues a wide variety of echolocation click-trains, it has not been found to emit whistles of any sort, and is regarded as incapable of doing so. Recent research suggests it is possible to relate specific sequences to individuals and behaviour by analysing the intervals between successive clicks in a sequence (Amundin, 1991). If the inter-click interval is converted to the pulse repetition frequency (PRF), it might be possible to use techniques developed originally to distinguish between frequency-modulated whistles but to apply them to a PRF-modulated click train from a harbour porpoise.

Automated Recognition Routines

The quantity of data accumulated from the Moray Firth trials was enormous. Cliff top observation started at approximately 8:30 a.m., and often did not end until diminishing light made sighting dolphins impossible at between 6:30 p.m. and 7 p.m. Recordings were made whenever dolphins were sighted in the area of the reflector barrier, which over the five weeks of field trials between 1991 and 1993 totalled over a hundred hours in duration. Spread throughout these tapes were the dolphin whistles that were to be analysed. Sometimes a group of dolphins would traverse the area entirely silently, or more commonly would emit only occasional whistles. Normally only the more active groups produced a deluge of vocalisations, but the position of each whistle on the tape needed to be identified and the whistle subsequently extracted.

Informal experiments on researchers in our group indicated a maximum hearing threshold of 14.5 kHz, and since bottlenose dolphin whistles can exceed 20 kHz then the tapes would have to be replayed at slower than normal speed if all whistles were to be detected by the human ear. During recording, it was noticed that a whistle would often begin at a low frequency within the human hearing range. Unfortunately one could not rely on this feature as a good detection system since one whistle that was analysed proved to contain only frequencies above 18 kHz — frequencies that are impossible to detect by most human ears. Listening to dozens of

tapes at half speed for whistles was not deemed to be a fruitful use of a researcher's time, and so an automated system of some description was required to detect the presence of a whistle on tape. This would allow the researcher to spend his time more profitably elsewhere.

Once a whistle has been detected on the tape the frequency modulations of the whistle are required. Sound sources are plentiful in the sea and it is far from a peaceful haven for the dolphins, only disturbed by the intrusion of unnatural man-made noise, as some of the more emotional admirers of cetaceans might suggest. Sound is used as a signalling system by a wide variety of animal life in the seas and oceans, including many species of crustacean and also several of fish. Additionally, sounds are produced by the actions of waves breaking, and cavitation from air entrained in the water, in addition to the action of currents moving pebbles or even small rocks against each other. Yet another unwanted contribution to all of this background noise is the dolphin itself. Echolocation clicks have a high intensity, much louder than that of the whistles albeit for a considerably shorter duration.

One of the commonest ways of representing sound signals is by use of a spectrogram, which shows a graph of time versus frequency with a changing intensity or colour on the graph representing different signal intensities. With all our preconceptions and experience of what we are looking for in the spectrogram, it is often still difficult to define precisely which parts are from the dolphin's whistle and which are merely noise. Taking into account the large number of whistles from the Moray Firth trials, some automated and preferably quantitative method is required to extract the time-frequency contour of the whistle from each spectrogram. Our thought processes in visually differentiating between whistle and noise need to be turned into algorithms that will perform the same task, but on a more objective level. More objectivity need not, however, be commensurate with increased validity of the results so obtained, and some human intervention (or at least double-checking) during this key activity may be required.

Dolphins have been shown to possess the ability to distinguish between signature whistles from different pool-mates (Caldwell *et al.*, 1972), so it appears reasonable to expect that there is some characteristic in the signature whistle that identifies that individual. Determination of the vocaliser of a whistle by comparison

of a whistle's frequency from instant to instant with some absolute frequency-time template held in the brain seems to be a rather complicated and cumbersome activity. When carrying out the same task with the aid of a spectrogram, the 'shape' of the whistle (i.e. the modulation of the whistle's fundamental frequency with time) appears to be the most obvious cue for discrimination. From this observation, any encoding of the whistle should preserve its acoustic 'shape' and reduce the significance of time-frequency components. This encoding should allow the use of a more efficient algorithm for identification of the signature whistle than if a comparison of the time-frequency sequence of the whistle signal itself were used, and additionally should not adversely affect the accuracy of the result.

The specification of a quantitative comparison requires that a probability be calculated of two whistles coming from the same animal or group of animals. In order for this to be achieved, a 'distance' measure between whistles is needed. If this distance measure has a well-known probability distribution, it can then be converted into a probability of the whistles having been produced by the same group or dolphin. By comparing a number of whistles between groups of animals, and assigning a confidence level at which two groups are taken to consist of the same dolphins, we may list the times at which the same group of animals passed through the trial area. This data may then be used in conjunction with the groups' movements, which were plotted with a theodolite, to study the dolphins' change in behaviour as they become habituated to the reflector barrier in the trials.

Once the initial aim of group identification is complete, the resulting analysis tool might be used to study other areas of interest in dolphin vocalisation. It might be possible to further determine signature whistle differences so that the same signature whistle voiced by different dolphins could be distinguished. This would allow identification of signature whistle mimicry of one dolphin by another. Dolphin vocalisations have been found to change with behaviour and stress (Caldwell *et al.*, 1990), so some information on the behaviour of whistling animals might be inferred using this analysis method. This may in turn give some measure of the level of arousal that is engendered by interaction with the reflector barrier.

Research Objectives

The primary reason for this research was the requirement to study changes in behaviour of wild dolphins for tests of acoustics net markers. The main obstacle to this study was the difficulty in identifying groups of animals. Photographic identification had not been utilised, for a number of practical reasons, although acoustic recordings were available for almost all the times that dolphins were observed in the area.

The main objective of this research is to quantify the probability that two groups of dolphins recorded acoustically during the sea trials were the same. These results can then be used to track the occurrence of specific groups of dolphins from day to day, and then aid in studies of the behavioural changes of the dolphins to objects placed in the water with them over multiple encounters.

Many tapes of these encounters have been made, and these must be examined for the required whistles. Since this is a very labour-intensive task, a more automated technique is desirable so that research time may be spent more profitably.

The individuality of the whistles seems to be connected with the frequency modulations contained within it, according to many previous researchers, and less to do with differences in intensity. A way of extracting and storing this information must be designed.

From this information, a database of whistles will be required against which candidate whistles may be compared. The probability that each whistle belongs to each class should be determined, in addition to the probability that the whistle belongs to a class not previously observed.

If this last objective can be achieved, determining the statistical significance of differences in the whistles can allow the completion of the prime objective, which is to determine the probability that two groups of dolphins are the same.

Thesis Outline

The tasks described under Research Objectives could be split into a number of sub-tasks. These have formed the outline for the thesis, and the majority fell into the following chronological order:

- Recordings of whistles from a number of sources in differing conditions were collected so that differences in whistles could be assessed (Chapter 2).

- Analysis tools were developed to aid in visualising the characteristic features of the whistles, and also to begin to study ways of extracting key features for classification (Chapter 3).

- From this initial data, filtering routines were created that reduced the background noise, and a whistle detection routine was developed to aid in determining where whistles were located on tapes (Chapter 4).

- Those frequencies contained with the dolphins' whistles were extracted from the other background noise in the signal, resulting in the whistles' 'contour' (Chapter 5).

- Characteristic features of the contours were assessed, and these were used to encode each contour into a more compact form suitable for the large comparisons needed for classification (Chapter 6).

- Studies of the differences in whistles between different individuals in addition to those from the same dolphin allowed syntactic and statistical pattern recognition techniques to be used to classify the whistle (Chapter 7).

- Analysis of groups of whistles from different times allowed a quantitative probability to be calculated for those whistles to have been produced by the same group of dolphins (Chapter 8).

A software package was developed that allowed most of the routines that were more often used together to exchange data with each other seamlessly. Although not described explicitly in any of the chapters, it contains the filtering, contour extraction, encoding, and classification routines. A paper describing the operation of the software package can be found in Appendix C.

References

Amundin, M. (1991). "Click repetition rate patterns in communicative sounds from the harbour porpoise, *Phocoena phocoena*." In *Sound Production in Odontocetes, with Emphasis on the Harbour Porpoise* Phocoena phocoena, PhD thesis, Department of Zoology, University of Stockholm, Sweden.

Bloom, P.R.S. (1991). "The diary of a wild, solitary, bottlenose dolphin (*Tursiops truncatus*), resident off Amble on the north Northumberland coast of England, from 1987 to January 1991." *Aquatic Mammals* **17**(3).

Bloom, P.R.S. (1992). "The movements and changing behaviour of a wild, solitary bottlenose dolphin (*Tursiops truncatus*) from January 1992 until his final disappearance in April 1992." Paper presented at European Association for Aquatic Mammals Symposium; Brugge, Belgium; March 1992.

Bloom, P.R.S., Goodson, A.D., Klinowska, M., and Sturtivant, C.R. (1995) "The activities of a wild solitary bottlenose dolphin (*Tursiops truncatus*)." *Aquatic Mammals* **21**(1): 19–42.

Caldwell, D.K. & Caldwell, M.C. (1977). "Cetaceans." In *How Animals Communicate* (T.A. Sebeok, ed.), pp. 794-808. Bloomington: Indiana University Press.

Caldwell, M.C. & Caldwell, D.K. (1965). "Individualized whistle contours in bottlenosed dolphins (*Tursiops truncatus*)." *Nature (London)* **207**: 434–35.

Caldwell, M.C. & Caldwell, D.K. (1971). "Statistical evidence for individual signature whistles in Pacific whitesided dolphins, *Lagenorhynchus obliquidens*". *Cetology* **3**: 9 pp.

Caldwell, M.C. & Caldwell, D.K. (1972). "Vocal mimicry in the whistle mode by an Atlantic bottlenosed dolphin." *Cetology* **9**: 1–8.

Caldwell, M.C., Caldwell, D.K., and Miller, J.F. (1973). "Statistical evidence for individual signature whistles in the spotted dolphin, *Stenella plagiodon*." *Cetology* **16**.

Caldwell, M.C., Caldwell, D.K., and Tyack, P.L. (1990). "Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin." In *The Bottlenose Dolphin* (S. Leatherwood and R.R. Reeves, eds.), Academic Press, pp. 199–234.

Caldwell, M.C., Hall, N.R., and Caldwell, D.K. (1972). "Ability of an Atlantic bottlenosed dolphin to discriminate between, and respond differentially to, whistles of eight conspecifics." *Proceedings of the Eighth Annual Conference on the Biological Sonar of Diving Mammals 1971*, Fremont, California, pp. 57–65.

Dahlheim, M.E. (1981). "Signature information in killer whale calls." *Whalewatcher* **15**(1): 12–13, 19.

Defran, R.H. & Pryor, K. (1980). "The Behavior and Training of Cetaceans in Captivity." In *Cetacean Behavior: Mechanisms and Functions* (L.M. Herman, ed.), pp. 319–362. Wiley (Interscience) New York. 463 pp.

Dreher, J.J. (1961). "Linguistic considerations of porpoise sounds." *Journal of the Acoustical Society of America* **33**(12): 1799–1800.

Dreher, J.J. & Evans, W.E. (1964). "Cetacean communication." In *Marine Bio-acoustics* (W.N. Tavolga, ed.), Vol. 1, pp. 373–93. Pergamon, New York. 413 pp.

Evans, W.E. & Prescott, J.H. (1962). "Observations of the sound production capabilities of the bottlenosed porpoise: A study of whistles and clicks." *Zoologica* **47**(3): 121–28.

Ford, J.K. & Fisher, H.D. (1981). "Killer whale (*Orcinus orca*) dialects as an indicator of stocks in British Columbia." *International Whaling Commission Scientific Committee Report* SC/JN81/KW8.

Ford, J.K. & Fisher, H.D. (1983). "Group specific dialects of killer whales (*Orcinus orca*) in British Columbia." In *Communication and behavior of whales* (R. Payne, ed.), AAAS selected Symposia Series, Westview, Boulder, Colorado, pp. 129-161.

Griffin, D.R. (1958). *Listening in the Dark*. Yale University Press.

Goodson, A.D. & Datta, S. (1992). "Acoustic detection of fishing nets: the dolphin's perspective." *Acoustic Letters* **16**(6).

Herman, L.M. (1975). "Inference and auditory short-term memory in the bottlenosed dolphin." *Animal Learning and Behavior* **3**: 43–48.

Herman, L.M., Richards, D.G., & Wolz, J.P. (1984). "Comprehension of sentences by bottlenosed dolphin." *Cognition* **16**: 129–219.

Herman, L.M. & Tavolga, W.N. (1980). "The communication systems of cetaceans." In *Cetacean Behavior: Mechanisms and Functions* (L.M. Herman, ed.), pp. 149–209. Wiley (Interscience) New York. 463 pp.

Kellogg, W.N. (1958). "Echo ranging in the porpoise." *Science* **128**: 982–88.

Kellogg, W.N. (1961). *Porpoises and Sonar*. University of Chicago Press, Chicago. xiv + 177pp.

Kellogg, W.N., Kohler, R., and Morris, H.N. (1953). "Porpoise sounds as sonar signals." *Science* **117**: 239–43.

Kole, K.R. & Speakman, J.R. (1993). "Measurement of the resting metabolic rate of the Atlantic bottlenose dolphin (*Tursiops truncatus*)." *European Research on Cetaceans* **7**: 247-50.

Lang, T.G. & Smith, H.A.P. (1965). "Communication between dolphins in separate tanks by means of an electronic acoustic link." *Science* **150**: 1839–44.

Lilly, J.C. & Miller, A.M. (1961). "Sounds emitted by the bottlenose dolphin." *Science* **133**: 1689–93.

Martin, A.R. (1990). *Whales and Dolphins*. Salamander Books Ltd., London and New York. 192 pp.

McBride, A.F. (1956). "Evidence for echolocation by cetaceans." *Deep Sea Research* **3**: 153–54.

McCowan, B. (1995). "A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, *Tursiops truncatus*)." *Ethology* **100**: 177–93.

McCowan, B. & Reiss, D. (1995). "Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins (Delphinidae, *Tursiops truncatus*): a re-evaluation of the signature whistle hypothesis." *Ethology* **100**: 194–209.

Penner, R.H. (1988). "Attention and detection in dolphin echolocation." In *Animal Sonar: Processes and Performance* (P.E. Nachtigall and P.W.B. Moore, eds.). Plenum Press, New York.

Reiss, D. & McCowan, B. (1993). "Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): evidence for vocal learning." *Journal of Comparative Psychology* **107**: 301–12.

Richards, D.G., Wolz, J.P., & Herman, L.M. (1984). "Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin (*Tursiops truncatus*)." *Journal of Comparative Psychology* **98**: 10–28.

Schevill, W.E. & Lawrence, B. (1953). "Food-finding by a captive porpoise (*Tursiops truncatus*)." *Breviora* **53**: 1–15.

Schevill, W.E. & Watkins, W.A. (1962). *Whale and Porpoise Voices*. Woods Hole Oceanographic Institute, Woods Hole, Massachusetts. Phonograph record and 24-page booklet.

Tavolga, W.N. (1968). "Marine animal data atlas." *Naval Training Device Center, Technical Report,* NAVTRADEVCEN 1212-2: i–x, 1–239.

Tyack, P.L. (1985). "An optical telemetry device to identify which dolphin produces a sound." *Journal of the Acoustical Society of America* **78**: 1892–95.

Tyack, P.L. (1986). "Whistle repertoires of two bottlenosed dolphins, *Tursiops truncatus*: mimicry of signature whistles?" *Behavioral Ecology and Sociobiology* **18**: 251–57.

Tyack, P.L (1989). "Use of a telemetry device to identify which dolphin produces a sound: When bottlenose dolphins are interacting they mimic each others signature whistles." In *Dolphin Societies: methods of study* (K. Pryor and K.S. Norris, eds.), University of California Press, Berkeley.

Wilson, B., Thompson, P., and Hammond, P. (1993). "An examination of the social structure of a resident group of bottle-nosed dolphins (*Tursiops truncatus*) in the Moray Firth, N.E. Scotland." *European Research on Cetaceans* **7**: 54–56.

Watkins, W.A. (1974). "Bandwidth limitations and analysis of cetacean sounds, with comments on 'Delphinid sonar measurements and analysis'." *Journal of the Acoustical Society of America* **55**: 849.

Wood, F.G., Jr. (1953). "Underwater sound production and concurrent behavior of captive porpoises, *Tursiops truncatus* and *Stenella plagiodon*" *Bull. Mar. Sci. Gulf and Caribbean* **3**: 120–33.

Wood, F.G., Jr. (1973). In *Marine Mammals and Man*. Luce, Washington, D.C. 264pp.

Chapter 2: Data Collection

Introduction

The data from the Moray Firth trials consisted of many tapes of recordings, but with no information about the identity of the vocalising animals. During any one recording period it can be stated that all the whistles belonged to animals in one group, but this does not give any information on how the signature whistles from one animal vary from one vocalisation to the next. If all signature whistles were repeated exactly, and whistles differed to a large extent from one dolphin to the next, then we would be able not only to differentiate between one group and another, but also to identify those times when one particular dolphin was present in the study area. At the other extreme, if a signature whistle varied greatly between vocalisations and all dolphins produced similar but slightly different whistles, we would be unable to use them to give us anything but the vaguest information on group identities, and certainly nothing about individual identity.

Clearly, the real world situation lies somewhere in between these two extremes, but it is not immediately apparent precisely where. If we are to calculate the probability that two whistles belong to the same 'class', be that an individual or a social group, we will need detailed information on the variation in signature whistles for any one animal. In the wild such opportunities where the vocalising animal can be identified are very rare, although they have occurred and have been studied (Dudzinski *et al.*, 1995). The identity of animals housed in an oceanarium, however, can be established more easily, and animals can be separated down to individuals if necessary so that there is no doubt about the source of any whistle. This isolation might lead to differences in the whistles produced by a dolphin under the stress of separation from its normal social group, so a more natural situation is to have a small group of dolphins that can interact normally with each other. A stereo pair of hydrophones and a video camera can be used to determine the source of most whistles in this situation, provided that the dolphins are not continually swimming in close proximity to each other.

By making a collection of whistles made by a single dolphin under a variety of conditions, the typical spread of its whistle characteristics can be determined. If this spread is applied to other whistles where the identity of the source animal is not known, the probability of any two whistles belonging to the same dolphin might be calculated. The assumption underpinning such a calculation is that any two dolphins will have similar variations in their whistle characteristics. This too can be determined from studies of the whistles from several dolphins when the whistler's identity is known.

This method of determining the spread of whistle attributes also facilitates the additional aims set out in the introduction. The determination of when mimicry is taking place can be achieved by identifying differences in the same signature whistle voiced by two dolphins if this behaviour is observed sufficiently often with dolphins in a closely studied situation, such as at an oceanarium. Changes in the whistles due to behaviour might also be quantified for use with wild populations providing a suitable method of classification of behaviour can be found.

Data Requirements

In order to quantify these variations, recordings of dolphins were made at three oceanariums — Kolmardens Djurpark, Sweden, and the Miami Seaquarium and Theater of the Sea, Florida. All recordings were made of bottlenose dolphins involved in a number of different activities. These were subjectively grouped into behavioural classes used by Defran and Pryor (Defran & Pryor, 1980):

- Affiliative/Social/Contact Behaviours.
- Aggression.
- Care-Giving.
- Fear/Stress/Subordination.
- Curiosity/Manipulation/Play.
- Sexual Behaviour.
- Leaping and Surface Behaviour.
- Training.

The final behavioural class 'Training' was not used by Defran and Pryor, but it was felt that grouping behaviours in evidence during a training or show session into, for example, 'Curiosity/Manipulation/Play' or 'Leaping and Surface Behaviour' could be misleading. The environmental factors that lead to a type of behaviour during a show may be very different from those in which the behaviour is offered spontaneously.

One of the assumptions made is that the behaviour that were observed in an oceanarium environment corresponded to similar behaviour in the wild. Certain pressure groups that oppose the keeping of dolphins in the artificial conditions of an oceanarium contend that this setting is so different from that of the wild that no research that relies on behavioural observations can be generalised to wild populations. One factor that makes this question difficult to answer is that, in general, wild dolphins react in a different way to the presence of humans than do those in human care who are habituated to their presence.

Lockyer (1990) studied incidents involving wild, sociable dolphins who have become familiar with having people in close proximity to them, as is the case with those dolphins in oceanariums. In answering the question "Is the behaviour of wild sociable dolphins similar to that of captive dolphins?" she states:

> "There is an interesting standard for comparison in Dolly, the ex-U.S. naval dolphin. She appeared to return satisfactorily to an independent existence in the Florida Keys, although she continued to seek human company. The types of behavior she exhibited have all been reported for captive dolphins, and the behavioral repertoire she displayed, once free, was similar to that of many truly wild sociable dolphins.
>
> ...
>
> The behavior of wild sociable dolphins thus appears very similar to that of captives, given the environmental restrictions of the latter." (Lockyer, 1990)

Having established that observations made of the behaviour of captive dolphins has validity when compared to the same behaviour in the wild, recordings were made at a number of oceanariums. The identity of the vocalising dolphin in the recordings was advantageous, but when this was not possible to obtain, the identity of the group of animals sufficed. Clearly information on individual identity could only

be obtained when an observer was close to the vocalising animal and was able to identify it.

The requirement for the recording system was that is should accurately reproduce the dolphins' whistles for analysis at a later time. The frequency response of the system was thus of importance, and a flat response from less than 3 kHz to more than 20 kHz was required, with a preferred frequency response of up to 30 kHz.

Oceanarium Recording Methods

In oceanariums with underwater viewing facilities, the identity of the vocalising animals is easier to establish by visual means. Additionally, a stereo pair of hydrophones (separated by roughly 75 cm), a video camera, and a hi-fidelity video recorder often allowed whistling individuals to be identified at the time of recording, and also later when the video tape was replayed. The facilities for doing this at the recording sites were only available at Kolmårdens Djurpark in their 'Dolphin Lagoon' area, but proved to be very useful (and were also very popular with the staff and public when stereo computer loudspeakers were connected to the output of the video recorder).

Kolmårdens Djurpark, Sweden

Initial recordings were made at Kolmardens Djurpark, Sweden, with an HS70 1" ball hydrophone, 20 dB low noise preamplifier, and a Nagra IV SJ recording at 7½ inches per second (ips). The recording level was adjusted in order to avoid keep the recorded signal within the amplitude range of the Nagra. The frequency range of the tape at 7½ ips on the Nagra was from less than 1 kHz to approximately 22 kHz, and the frequency response of the hydrophone was nominally flat over this range

Three bottlenose dolphins were contained in the pool, which were all females formerly cared for at Flamingo Land Dolphin Centre, North Yorkshire. The animals had been housed together for many years, and had formed what Peter Bloom, their owner, considered to be a stable social group.

The Dolphin Lagoon consisted of a large, irregularly shaped pool with a depth varying from less than 1 metre up to 9 metres, surrounded on one side by a series of

underwater viewing windows of approximately 3 metres height. The presence of these windows, and the lack of good visibility from other positions around the pool, suggested a position at one end of the windows was a good place from which to make recordings. This location lay in the corner of the pool next to a channel connecting to the other pools in the complex, and reduced the contribution from wall reflections unless the dolphin was in the channel itself.

Recordings were made on a single channel of the Nagra during echolocation trials with the dolphins, and also during periods of activity outside the trail times. Notable behavioural events and instances in which one dolphin was identified as the vocaliser were indicated by a series of tones placed on the tape, along with a written log kept of identities, times, and tone sequences. Peter Bloom made the identification of the whistling individual when any loud whistle was heard simultaneously with bubble streams observed from an animal's blowhole. Although this identification technique is open to error, when a loud whistle is observed with only a single dolphin emitting a bubble stream, it was adjudged highly probable that the vocaliser and the animal producing the bubbles were the same individual.

A similar recording technique was used with the substitution of an Aiwa Digital Audio Tape (DAT) HD-S1 recorder in place of the Nagra on two subsequent occasions at Kolmården, in March and May 1995. The frequency range of the DAT recorder is limited to 22 kHz (for a 44.1 kHz sample rate DAT). From analysis of the recordings made by Santhi Mahadevan (Mahadevan, 1994) and early recordings made at this facility with the Nagra, this recording range appeared to cover the normal range of frequencies employed by these dolphins. Additional recordings also were made of the other dolphins at Kolmården at this time, and concurrent identification of the vocalising animals by the trainers was recorded on the second channel of the tape.

Recordings of the dolphins at Kolmården were made in as many different conditions as was practical, although recordings during public shows in the show pool were not practical since there were no offstage areas sufficiently close to the pool. Training sessions were recorded, however, as were also the short shows presented in the 'Dolphin Lagoon', which is the Park's dolphin breeding pool. In addition to the three ex-Flamingo Land dolphins, whistles were recorded from five adult females and

one adult male dolphin, and one juvenile (two year-old) female and three neonate dolphins (one female, two males).

During May 1995, the seven days old calf of one of the female dolphins in the Lagoon area (Nephele) died mid-morning, and she continued to emit loud and frequent signature whistles for the rest of the day. These were recorded and taken as indicators of whistles produced under very stressful conditions. No other whistles from similar stressful circumstances were obtained during the study period. Unfortunately, Nephele was one of the less vocal dolphins, possibly due to her having moved only recently to the facility with Delphi from Germany in December 1994. This lack of other baseline vocalisations makes it difficult to make best use of this piece of data, which proved to be unique in our oceanarium recordings.

In order to capture changes in whistles with behaviour, in November 1995 recordings were made using a stereo pair of hydrophones placed in the Dolphin Lagoon. These were connected to a 20 dB low noise four-channel pre-amplifier (only two channels of which were used) and thereby to a NICAM stereo VHS video recorder. A Hi-8 CCD video camera was positioned to observe the activity of the dolphins through the underwater viewing windows, and this provided the video signal that was recorded in addition to the hydrophone audio on the video recorder. The headphone output from the video player was connected to two active speakers (of the sort commonly used with computer sound cards), which provided an excellent method of monitoring the acoustic activity that was occurring in the pool. If the speakers were placed too close to the pool, however, and adjusted so they were too loud, then the sound would propagate back into the pool and attract the dolphins' attention, thereby disrupting the behaviour that we were trying to observe.

Four dolphins were present in the Dolphin Lagoon during the simultaneous video and acoustic recording — three adult females, and one neonate male (5 months old). Seven hours of video tapes were recorded from a variety of behaviours presented by the dolphins, although most vocal activity seemed to take place during shows and training sessions.

Miami Seaquarium & Theater of the Sea, Florida

The same DAT based equipment as was used to record the dolphins in Sweden was also used to record bottlenose dolphins at the Miami Seaquarium in their 'Top Deck' dolphin pool, and in the 'Flipper Stadium'. The Top Deck pool consists of a large circular pool that has been in existence at the Seaquarium for many years, and is approximately 6 metres deep and 15 metres in diameter. The area was in use as a retirement and nursery facility for the dolphins, and it was easy for the trainers to keep the pool under observation. One of the younger dolphins in the Top Deck pool took a great interest in the hydrophone, which made it necessary to withdraw the hydrophone from the water when he approached it too closely.

One of the difficulties experienced when high value equipment is used near animals as curious and capricious as dolphins is the risk of incidental damage to the equipment. In this case, the dolphin would approach the hydrophone, echolocating on it, and then try to grab the hydrophone and swim off with it. If he had succeeded, the hydrophone, pre-amplifier, and possibly even the DAT recorder might have ended up in the pool. Salt water and dolphins' teeth do not form the optimal environment for the survival of electronic equipment from the researcher's point of view, and neither does electronic equipment small enough to be swallowed provide the optimal toy for a dolphin from the trainer's point of view.

When considering data collection from dolphins, especially those that have become habituated to humans and all their associated paraphernalia, safety of both the equipment and dolphins must be given a high priority. Dolphins have a predisposition for soliciting reactions to their presence from humans, and then to attempt to control the situation, and pulling on the end of a hydrophone cable seems to satisfy both these needs. It is important that, for example, the combination of withdrawing of the hydrophone from the water combined with the dolphin's attempts to grab it before this occurs, should not turn into a game of tactics and speed of reactions, as disaster will most assuredly follow. Some knowledge of dolphin behaviour is desirable in these situations, and, if one is insufficiently familiar with this, a trainer should be present who can notice the precursors to 'game playing' behaviour that might in turn lead to damage to equipment or dolphin. The best way to avoid this sort of situation is to 'be boring', by quietly watching the dolphin but not talking to nor waving your arms at it.

Any equipment should be withdrawn from the pool at the first sign of unwanted interest, and then it should be replaced only when it is clear that the dolphin has *actually* (rather than just apparently) lost interest. Dropping hydrophones into the water with a satisfying 'plop' is a very efficient way of attracting a dolphin.

The recordings at the Top Deck site were made from the side of the pool closest to the trainers' area, this being the only location with a power point nearby. Unfortunately, the dolphins were used to being reinforced with food from this area, and consequently seemed to view the hydrophone with a higher level of curiosity than might have been evidenced from other places around the pool. The staff suggested a solution to this unwanted curiosity, and they offered to conduct a training session at one of the platforms some distance away from the recording point. By the time the training had finished, most of the dolphins seemed to have accepted the presence of the hydrophone.

The 'Flipper Show' area consisted of a roughly rectangular pool excavated from the natural rock, which had been used during filming of the original 'Flipper' series. The water was shallower than that in the Top Deck pool, averaging near 3 to 4 metres, but the surface area was far greater (estimated at 30 metres by 20 metres). Recording took place from a holding pool made of wooden bars spaced sufficiently far apart such that sounds from nearby areas were not obscured. The gate to the holding pool was closed, blocking the dolphins' access to the hydrophone and thereby protecting the equipment. The dolphins often frequented a pool close to the holding pen outside show times, and the majority of whistles recorded from this area were tentatively placed in the 'Affiliative/Social/Contact' class of behaviours.

Recordings made at Theater of the Sea, Islamorada, used the facility's own hydrophone and preamplifier system, since access to power outlets close to the water was not available. The main enclosure consisted of a long stretch (estimated at 80 m) of natural seawater surrounded by land on three sides, and a chain-link fence on the seaward side. Smaller pools were fenced off to allow temporary separation of the dolphins for husbandry reasons. The water depth varied considerably, and no information was available on the bottom profile.

The recording position was located next to the main stage that used during the dolphin presentation. Unfortunately, the power supply to the DAT recorder seemed to

interfere with the radio microphone receiver, and so no recordings were made during shows.

Recording Methods at Sea

The data requirements for the recordings made during the field trials in the Moray Firth were similar to those for oceanarium recordings, except in this case it was initially impossible to identify the whistling animals or the groups as they were recorded. The frequency response necessary for the recording system was the same, from 3 kHz to 20 kHz, and with the sensitivity as flat as possible.

Recordings were often made from sonobuoys, and when the batteries in these were reaching the end of their charge, the signal to noise ratio would decrease and resonances would start that were typically around 10 kHz, which would change the frequency response. Directly cabled hydrophones were available on the last of the three trials, which didn't suffer from the same battery lifetime limitations. Equipment details on each of the trials have been reported in the literature (Goodson *et al.*, 1994; Klinowska *et al.*, 1992; Mayo & Goodson, 1992; Woodward *et al.*, 1993).

One of the most intensively studied periods was for interactions from 14th to 16th September 1992. The logs from these days have been studied, and all the whistles from times when dolphins were in the area have previously been sampled as part of a cataloguing project. These whistles formed one of the test data sets for assessing the applicability of the final classification system to practical field data. Eleven different interactions were identified where one or more whistles were recorded. Clearly, when dolphins traversed the trial area without emitting any whistles it was not possible to use this method for identifying them, and these interactions were not included in the test data.

Summary

Initial recordings from the Moray Firth trials contained a large number of whistles, but with no information on the identity of any of the vocalising animals. Further recordings made with dolphins in oceanariums took place either with a small, known group of animals, or where the vocalising dolphin could be identified. These

recordings provided data for calculating the spread of signature whistle attributes, both within a group, where mimicry of a whistle may occur, and also for signature whistles from one individual.

The recordings contained the whistles to be studied, but these were scattered irregularly throughout the tapes. Manually identifying portions of the recordings that contained a whistle would have been most time consuming, especially for the tapes made during the Moray Firth field trials. Some automated method was required to locate where the whistles were located on the tapes, and to sample them into a computer for analysis.

References

Defran, R.H. & Pryor, K. (1980). "The Behavior and Training of Cetaceans in Captivity." In *Cetacean Behavior: Mechanisms and Functions* (L.M. Herman, ed.), pp. 319–362. Wiley (Interscience) New York. 463 pp.

Dudzinski, K.M., Clark, C.W., Würsig, B. (1995). "A mobile video/acoustic system for simultaneous underwater recording of dolphin interactions." *Aquatic Mammals* 21(3): 187–93.

Goodson, A.D., Mayo, R.H., Klinowska, M. and Bloom, P.R.S. (1994). "Field testing passive acoustic devices designed to reduce the entanglement of small cetaceans in fishing gear." In *Cetaceans And Gillnets, IWC Special Issue* 15:597–606.

Klinowska, M., Goodson, A.D., Mayo, R.H., & Bloom, P.R.S. (1992). "Progress in the development of warning devices to prevent the entrapment of cetaceans (Dolphins, Porpoises & Whales) in fishing nets." ICES CM, ICES Conference, Germany.

Lockyer, C. (1990). "Review of incidents involving wild, sociable dolphins, worldwide." In *The Bottlenose Dolphin* (S. Leatherwood and R.R. Reeves, eds.), Academic Press, pp. 337–53.

Mahadevan, S. (1994). "Identification of individual bottlenose dolphins by analysing their signature whistles." Masters Thesis, Electronic and Electrical Engineering Department, Loughborough University, England. 124 pp.

Mayo, R.H. & Goodson, A.D. (1992). "Interactions between wild dolphins and a moored barrier: initial results from the 1991 Moray Firth Trail." *European Research on Cetaceans* 6.

Woodward, B., Morphett, N., and Goodson, A.D. (1993). "Tracking cetaceans in the vicinity of fishing nets." Commission of the European Communities, Study Contract Report 92/15.

Chapter 3: Preliminary Whistle Analysis

Introduction

An automated system was needed to analyse the large amount of data from the recordings made during the trials in the Moray Firth. The recordings consisted of large sections either of echolocation signals or no dolphin vocalisations whatsoever, with periods of whistle activity interspersed. Three broad tasks could be outlined that needed to be accomplished to determine the identity of each group:

1. Determination of the presence of a whistle, and isolation for analysis.

2. Extraction of the whistle's characteristics from the signal.

3. Comparison and classification with a previously held database of whistles.

Working within the mathematical realm of raw signals, in which analysis is conducted on the changes in voltage returned from the hydrophone and recorded on to tape, can be quite daunting. The more favoured method, which has been used to study dolphin vocalisations for many years, is to view the signal in the form of a spectrogram. Time and frequency axes are used together with changes in colour or shading to indicate the energy contained in different frequency bands during each time period (see Figure 1-2, first chapter for an example). This approach transfers the task of analysis of the whistle from one of sound analysis to that of image analysis, which is intuitively a simpler problem to grasp. Dolphins live in a world where sounds often take on more significance than images, but humans are still predominantly visually based animals.

The problem that initially presented itself was that of how to separate the whistles from the other sounds that had been recorded, when one knew it was contained in a sample. This would lead onto a method for determining the presence of a whistle on a recording, which would immediately produce a reduction in the amount of time spent on unwanted data.

Coding Strategy

It was acknowledged from the start that the design and implementation of a software package to identify, filter, extract, encode, and then classify whistles was going to require several thousand lines of software code. One of the main problems when writing a large (i.e. more than 5,000 lines) software application, is that of identifying and correcting programming errors. If one is realistic about the period of time that should be spent in designing an application of this sort of complexity, it is virtually impossible to write it without errors occurring when it is first executed. Errors can be due to typing mistakes, misuse of variables, unintended side-effects of code on other sections of the application, and also errors in the design itself. There are, fortunately, a number of methods to minimise these risks, and thereby to reduce the time that must be spent in 'debugging' the application.

Coding Without Consternation

Modular design of the application is a technique that ensures that each individual task is carried out in a self-contained way and with its access to the rest of the application under strict control, whether this is internal access to other modules or to external data. A second technique, which at first might seem obvious, is to spend time on overall design of the application. Detailed design is not always possible when the extent of the application is not known at the start of the project, but it is counterproductive immediately to leap in and start writing code before one has a strategy for incorporating it with other parts of the application. For instance, an application to calculate the frequency spectrum of a signal might return the colours that should be printed on the screen for the spectrogram, but this is quite useless if any further processing is required on the frequency spectrum itself. Good design allows the same piece of code to be used by many other routines, avoiding the necessity of writing many code segments that differ from each other just by a few lines.

The most important aid in correcting errors in a program is to provide adequate documentation. Design notes must be kept and updated as the code is modified, rather than leaving a stack of paper with different strategies for attacking a problem where it is not clear which of these have been superseded. The easiest and often most effective way to document code is by inserting comments at the end of every line, which briefly

explains its function. A longer description at the top of each module describes its function, what data it requires, and what it will return to the calling procedure. Without these aids it is surprisingly easy to lose track of exactly what a particular routine was designed to accomplish. A brief note in the code can save many hours of frustration and reverse engineering.

All of these methods were employed to minimise the possibility of errors in the software. The final integrated analysis application contained over 30,000 lines of code, and although there may still be minor problems, the majority of those that did occur during development could be corrected within half an hour. The longest problem to overcome took five days, which was a problem with two methods of memory allocation in use by the compiler, but due to the modular design and well commented code even such a system problem was isolated and corrected.

Development Environment

A number of options was considered concerning which programming language to use, but the language finally chosen for the project was Kernighan and Ritchie's 'C' language (Kernighan & Ritchie, 1988). Borland 'Turbo C', a 16-bit compiler, was initially used and is a commonly used compiler, which in addition is readily available on the university's IBM PCs. Later, a 32-bit compiler was used which gave significant increases in both speed and memory access (the Free Software Foundation's Gnu CC compiler, ported to the IBM by D.J. Delorie).

There were several advantages to using 'C' that influenced its choice. These advantages included low level access to hardware such as a sound card, flexibility of data structures and access, easy file access, and a broad knowledge base of programming techniques and pitfalls from several years of programming in 'C' on UNIX workstations. However, potential disadvantages included the difficulty of interpreting undocumented code, programming errors producing seemingly inexplicable machine problems due to unintended low level machine access, and the fact that 'C' has few advanced features for modular programming.

It was decided that conversion and analysis routines would be written as separate programmes rather than incorporated into one large application. This would allow as much flexibility as possible in the development of the software, and allow

programs to be modular. For this approach to be feasible, data files were used to pass information between programs, with library routines allowing data to be read from and written to these formats.

The programs that were written to perform the tasks shown in this section could be combined to form a research tool useful in a variety of applications. No special hardware was required to run the routines, and signal input and output could be performed by standard 16-bit sound cards, which are becoming widely available for the IBM PC. However, the test signals used for development of the routines were stored internally as files, and not directly sampled from tapes by the software.

Spectrograms and the Fourier Transform

The spectrogram is a commonly used technique to visualise sounds, and is a useful tool for acoustic analysis (Figure 1-2). The signal is converted from an amplitude-time domain to that of amplitude-frequency. Splitting the signal into segments over which the spectrum is calculated provides the time varying structure of the spectrum. The time interval over which the spectrum is calculated has an important effect on the spectrogram. Clearly, if the time partition (i.e. period) is long, very short impulsive signals will have little effect on the spectrum when compared to the rest of the signal in the partition. Conversely, if a short time partition is used, then short term signals will have a relatively greater effect than previously, albeit at the expense of resolution on longer period variations.

It is normal to convert the amplitude component of the spectrum into equivalent power (proportional to the square of the amplitude) and then to use a logarithmic scale to display the power spectrum. Differences in the power spectrum can be indicated using shades of grey or by using colour changes. Ideally, both the absolute power level and small differences to a value nearby should be easily identifiable from the display. In practice these two features form a trade-off, and using a cursor to interrogate power intensities on screen forms an acceptable compromise for determining absolute values.

35

The Fourier transform is normally used to calculate the frequency spectrum from the original signal. The equation for transforming between amplitude-time values and amplitude-frequency values is:

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{2\pi i f t} dt \quad \dotsfill (2\text{-}1)$$

where $h(t)$ is the original signal (amplitude changing with time), and $H(f)$ is the transformed signal (amplitude changing with frequency). It can be seen from the equation that the transformed function is in the complex domain, such that the amplitude at each frequency is represented by a real and an imaginary part. This two-dimensional component is due to the frequency having both phase and amplitude components. When calculating the power spectrum, since the square of the amplitude is taken, the complex amplitude becomes a wholly real power.

The signals that we manipulate with the computer, however, do not have the continuous nature suggested by the equation. The amplitudes are recorded at a regular time interval specified as a sample rate, which therefore has a specific effect on the frequency range that can be calculated. If one imagines a signal with a frequency f which is sampled at a rate of $2f$, then the sampled points could indicate a signal with any amplitude value between zero and the full amplitude of the signal. There is therefore an upper limit to the frequencies that can be represented correctly by a sampled signal, termed the Nyquist frequency, which is half of the sample rate. The Fourier transform of a signal which consists solely of real values (as opposed to a complex signal) is symmetrical, and so all values from 0 to $\frac{1}{2}f$ are repeated from $\frac{1}{2}f$ to f, and correspond to the negative frequencies $-\frac{1}{2}f$ to 0. It can therefore be seen that the Fourier transform of a real signal only specifies values for up to half the sampling rate of the original signal.

The 'fast' Fourier transform (or FFT) algorithm was used to convert the signals into a power spectrum, which as its name suggests is simply a fast way of calculating the Fourier transform of a signal. An explanation of the Fourier transform and its discrete equivalent, together with the derivation of the FFT algorithm, is given in Appendix B.1. (See also Press *et al.*, 1986.)

The readsig() Viewer

A stand-alone programme named readsig() was designed to view the signal after transformation by the FFT, and to save the resultant FFT data to disk for use by other programmes. The method used to display the power spectrum, however, requires further explanation.

The colour palette used to display the power levels for the FFT was chosen to maximise three criteria:

1. Ease of identification of absolute power levels;

2. Ease of differentiation between similar power levels;

3. The ability to identify the higher of two similar power levels.

A colour palette was chosen based on the HSL (hue, saturation, luminance) colour model with low power values indicated by a blue hue and low luminance, which extended through green, yellow, pink, and finally white with linearly increasing luminance for higher power values. Details of the options considered and the reason why this (non-standard) colour palette was chosen are given in Appendix B.2.

Transform Display

The transform display software was written so equal periods of time would be displayed as equal distances along the x-axis on the screen. Thus, transform partitions containing a larger number of samples will appear wider on the display to indicate that they are taken over a longer period of time. Vertically, all of the transformed values will be displayed on the screen, which will be values for all frequencies up to half of the original sampling rate of the signal.

In this way, no matter the size of the Fourier transform partition, a signal with a three second duration recorded at a sampling rate of 40 kHz, for example, would always appear with the same width and have a frequency scale from 0 to 20 kHz vertically. The effect of reducing the size of the partition would be to decrease the number of frequency 'bins' vertically, and to increase the number of time 'partitions' horizontally, resulting in blocks that are thinner but consequently taller. The converse would obviously occur if the partition size were to be increased. Figure 3–1 shows two example spectrograms from this program.
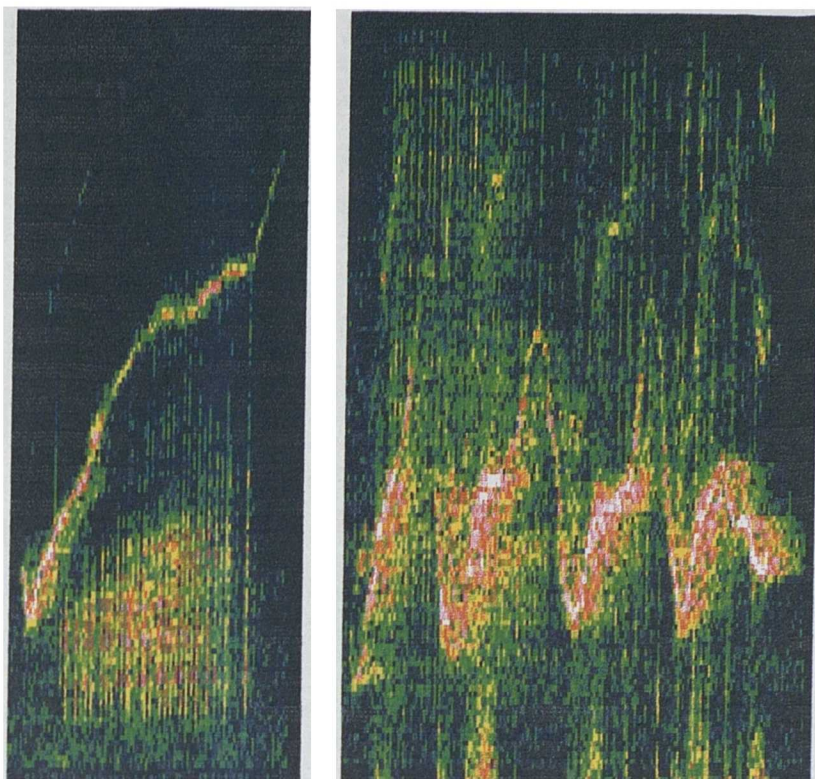
Figure 3-1: Two spectrograms from the 'readsig' program

Available Input File Formats

The readsig() programme could read either 16-bit signed integer files (here termed 'raw' data files), in which case the programme would calculate its Fourier transform, or it could read in previously saved FFT files, which are identified by the suffix '.FFT'. If a raw data file were read in, then the programme prompts for the size of the Fourier transform to use (expressed in 'number of bits') and would ask whether the results should be sent to an FFT data file. Since this implementation of the FFT routine required the partition size to be a power of two, using n 'number of bits' would result in a partition that is 2^n samples wide. The resulting FFT could be saved to disk for use by other programmes. In this case the default FFT file name would have the same body as the input file name but with the suffix '.FFT'.

When raw data was used as input to the routine, changes in the recording and playback levels of the original signal often required alterations to the scale used for the displayed power spectrum. To obviate this need, the maximum amplitude of the signal was determined and used as a scaling factor in order to normalise all data values to the range ± 1.0.

38

3d() — Viewing In Three Dimensions

Although the viewer in the readsig() programme allowed good differentiation between intensities, it was not always possible to distinguish the relative intensities between, for instance, two whistles, or a whistle and clicks in a click train. In order better to appreciate the structure of the data so that algorithms could be designed to extract features from it, a programme which provided a three-dimensional view of the FFT data was developed, named the 3d() routine.

This method transformed the power for each frequency bin into a 'height' (z-axis) on a three-dimensional map by converting the power into decibels and including a constant offset in order to raise peak values above zero. The x-axis was encoded as time, and the y-axis as frequency. An oblique projection was then used to place the y-axis at approximately 60° to the plane of the screen and away from the viewer, and a shear of 30° along the x-axis was introduced. Although this co-ordinate system was no longer orthogonal, the resulting image appeared natural to the eye, and included many of the desirable features of a full three-dimensional viewing system without the same computational overheads inherent in using a full matrix transformation routine.

An artificial 'floor' was placed on the data at drawing time whereby any data point that had a 'height' value less than zero was set to zero. This allowed the main features of the signal to be visible without confusion from low-level background 'noise' in the form of an irregular bottom to the surface. The array of points was mapped as a triangular mesh. After transforming the points according to the oblique projection parameters, the surface normal to each triangular facet was calculated (see Appendix B.3 for implementation details).

Four base colours were used to indicate the height of the data points in each triangle. The maximum height value was calculated from the three triangle vertices, then if this height was not greater than zero, the colour was set to blue. Colours of green, orange, and white were used to indicate ascending heights above zero. If these colours had been used with no further modification, however, it would have proved difficult to determine the boundary between adjacent triangles of the same colour. For this reason a lighting model was used to calculate a 'brightness' value for each

polygon based on the ambient lighting and diffuse reflection formulae. Figure 3–2 shows example output from the 3D program.
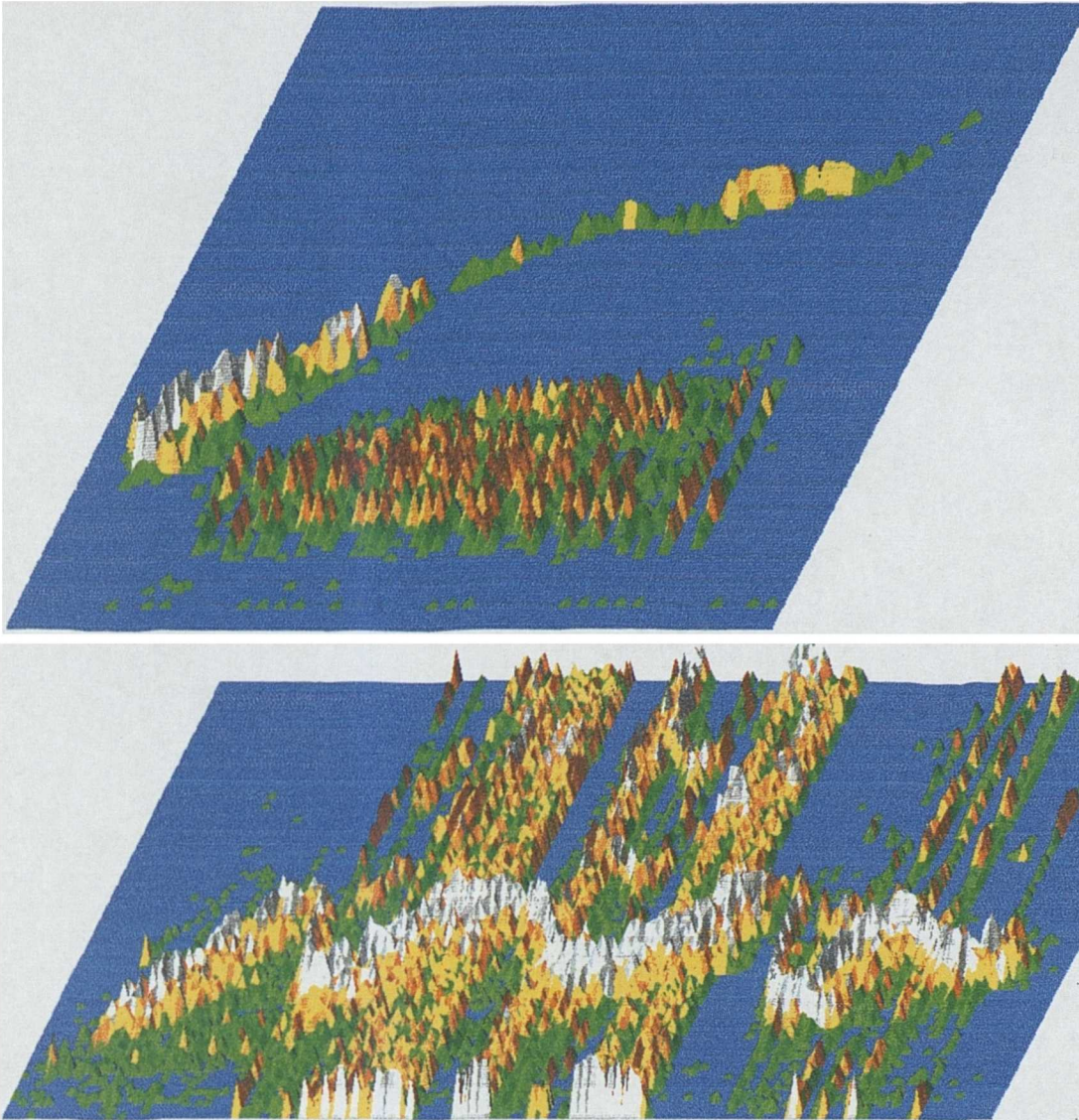
Figure 3-2: Example output from the 3D spectrogram program (time increases left to right)

Whistles and Trees

Initially it was proposed to construct the minimum spanning tree (MST) of the data points in the Fourier transform (see Diday and Simon, 1976), using the geometric distance between points as the weighting for the edges between nodes. The 'height' metaphor for intensity that was introduced in the 3d() programme was to be used to accomplish this task. It was envisaged that the resultant structure would provide a powerful way to extract information on both whistles and clicks, since these would

form chains of nodes separated from their surroundings by a highly weighted (large distance) edge.

The calculation of the MST, however, proved computationally very expensive, and in addition was costly due to the amount of storage needed to accomplish the task. When running on a 50 MHz 80486 DX processor, with programming access to 1 Mb of RAM for storage, the programme took just under 40 minutes to calculate the MST for around 25% of the data in a 128 x 127 Fourier transform array before running out of storage capacity. It was the long time period required for calculation of the MST that led to the decision to find other less computationally expensive ways of extracting information from the FFT data. Some time and effort went into the construction of the MST algorithm, however, and since it might prove useful for future applications, a description of the algorithm itself is given in Appendix B.4. Later a faster version of the 'C' compiler became available, which made use of 32-bit instructions and data access rather than the previous 16-bit limit, and had extended access to all available memory on the computer as well as up to 1 Gb of virtual memory. With this additional capacity and speed, it is possible that MST routines might again prove worthy of consideration.

Discussion of the Analysis Techniques

The routines used to analyse the whistle signals were all programmed in Borland Turbo C, a commonly used programming language that includes graphical extensions. The code produced by this compiler used 16-bit instructions and data access, however, whereas the 80486 and Pentium processors increasingly in use today can operate using 32-bit instructions. This allows them to make use of all physical memory on the machine as well as virtual memory without the need to write complex and time-consuming drivers.

Routines such as the minimum spanning tree encountered problems related to computation time and memory size restrictions. Once a suitable 32-bit compiler had eventually been obtained, an adequate contour extraction process had already been found. If the extra speed and memory made available by this compiler had been available earlier, however, a different (and possibly more efficient) solution might have been achieved.

Visualisation of the whistle signal spectrum with the three-dimensional display was considered to be the easiest way of interpreting the spectrum information, although the process could be quite slow since more calculation had to be done for the illumination and viewing model than for the standard two-dimensional spectrogram.

Routine modularization by the use of individual programs and data files for the transfer of data appeared to work successfully. From a practical point of view, however, it was often irksome to be obliged to terminate one program, type in a new command line to start up the next and wait for it to load, then terminate that program and move on to the next. Since many of the programs were rather voracious in their memory requirements within the 1 Mb limit imposed by the compiler, this modularization into separate programs was essential. After the 32-bit compiler allowed easy access to a virtual memory environment that was larger than the physical memory in the computer, this restriction was removed.

The more frequently used of the initial analysis routines were combined into a suite of utilities, filters, and displays, and translated to work under the GNU free software foundation's 32-bit GNU CC compiler.

Summary

Conversion of the whistle signal from a stream of amplitude-time data to an amplitude-frequency-time representation enabled the whistles to be visualised in a much more natural and intuitive way. Once in this form, three different methods of representation were considered with varying degrees of success, each suggesting a different way in which the whistle's frequency-time contour could be extracted from the background noise.

The first method was to represent the data in a standard 'spectrogram' form, with different shades and intensities of colour used to identify the power contained in the signal for each frequency and time partition. Advantages to this method were low computational overheads, and a commonly used display that users familiar with the field of acoustic analysis could easily assimilate. Two of the disadvantages were the difficulty in judging relative intensity values, and also that there was no obvious way that the whistle data could be isolated from noise components.

A method of representing the power levels as peaks in a three-dimensional landscape gave a more intuitive view of the signal. The whistle peaks generally showed themselves more easily than for the standard spectrogram, and the use of colour to indicate relative power levels and shading from a diffuse lighting model enabled a great deal of data to be presented in a clear and uncluttered way. A disadvantage was the time taken to calculate the three-dimensional viewing co-ordinates and data for the lighting model. Although visually dissimilar to the spectrogram, the internal data representation of the whistle signal was the same and so did not aid in whistle contour isolation.

The use of a minimum spanning tree connecting points in the amplitude-frequency-time three-dimensional plane looked initially to be a very promising representation. This method produced a data structure that was very close to the required separation of whistle and background, and only relatively simple procedures would have been required to accomplish this aim. Problems with storage space and calculation times for this representation for anything but very short signals precluded it from being a practical analysis tool, however.

By examining the metaphor of a three-dimensional landscape, techniques were envisaged for reducing background noise and enhancing the whistle component. Since these could be achieved with either the three-dimensional representation or the standard two-dimensional spectrogram, the spectrogram was chosen as the preferred representation due to its speed of computation.

The next task to consider was the detection of the whistle on the recordings, which also encompassed techniques for reducing the contributions from background noise in the signal.

References

Kernighan, B.W. & Ritchie, D.M. (1988). *The C Programming Language*. Prentice-Hall, Englewood Cliffs, new Jersey.

Diday, E. and Simon, J.C. (1976). "Clustering analysis." In *"Digital Pattern Recognition"*, K.S. Fu (ed.), Springer-Verlag, New York.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1986). *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, Chapter 12 Fourier Transform Spectral Methods, pp. 390–395.

Chapter 4: Whistle Enhancement and Detection

Introduction

It became clear that some form of automatic whistle detection routine would be needed to search through the Moray Firth trial recordings. The human ear is not designed to receive the high frequencies contained in dolphin whistles, and there was insufficient time or inclination to replay all of the tapes at speeds where the whistles would fall into the audible range.

The problem of whistle detection is related to that of isolation of the whistle's time-frequency contour, but in a less precise manner. In the case of detection, simply the times when whistles are present is required, and no information on frequency or other whistle attributes need be stored. The relatively long and unchanging frequencies contained in the whistle's spectrum could be seen quite clearly with the three-dimensional viewer that was described in the previous chapter, and these features seemed to hold the answer to differentiating whistles from background noises.

Snaps, Clicks, and Breaking Waves

The open sea is far from a quiet place, and provides many barriers to detecting the dolphins' whistles. One must try to minimise the number of 'false alarms', which are caused by wrongly identifying sounds from other sources as being a whistle. Apart from the whistles themselves, the sounds most apparent on the recordings came from one of three sources:

- other animal life, such as snapping caused by various crustaceans;

- the dolphins' broad-band echolocation click trains, which often could be seen as regular vertical bars of the spectrogram;

- other sounds from the environment, such as wave noise and pebbles moving against each other.
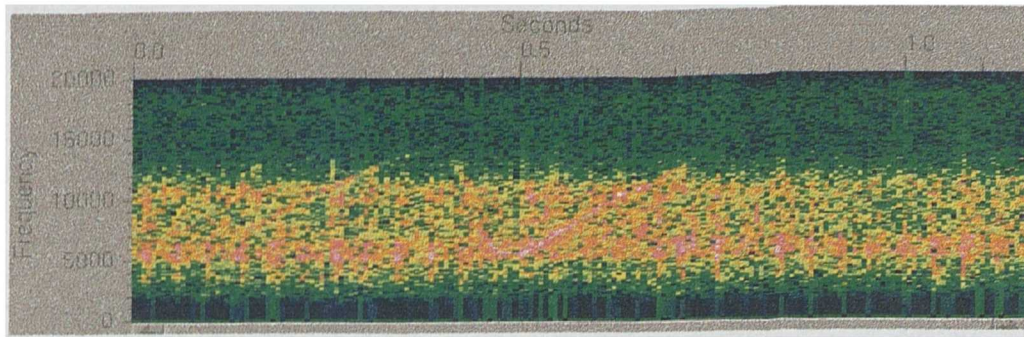
Figure 4-1: Spectrogram of two whistles obscured in noise

The features that made the whistles conspicuous on the displays were their long time duration, their narrow frequency bandwidth at any one instant, and their relatively slow change in dominant frequency. Snaps and echolocation clicks tended to have a very short duration and a wide frequency spectrum. Wave noise had slightly different characteristics, being longer than a click but still exhibiting a wide spectrum. The signal to noise of the whistles was often rather poor, especially when the vocalising animal was at a distance from the receiving hydrophone. In order to detect the whistle's presence in this background of wide-band and impulsive noises, it was decided firstly to increase the signal to noise ratio of the whistles.

Whistle Enhancement

After studying the displays from the three-dimensional viewing programme, one of the main obstacles to identification of the whistles appeared to be the contribution made by impulsive sounds, mainly attributable to echolocation clicks from the dolphins. These clicks were characterised by their short duration and their broad, predominantly uniform spectrum through the 0-22 kHz region. Additionally, if methods could be developed to boost the energy contained in those parts of the signal from the whistle, then their detection would greatly be simplified. The problems inherent in whistle detection are demonstrated in Figure 4-1, which contains two whistles hidden by background noise.

Reducing Background Noise

The broad spectral nature of the echolocation clicks especially, but also much of the other background noise, led to an initial attempt at noise reduction based upon this difference from the whistle component. Each individual pulse in an echolocation
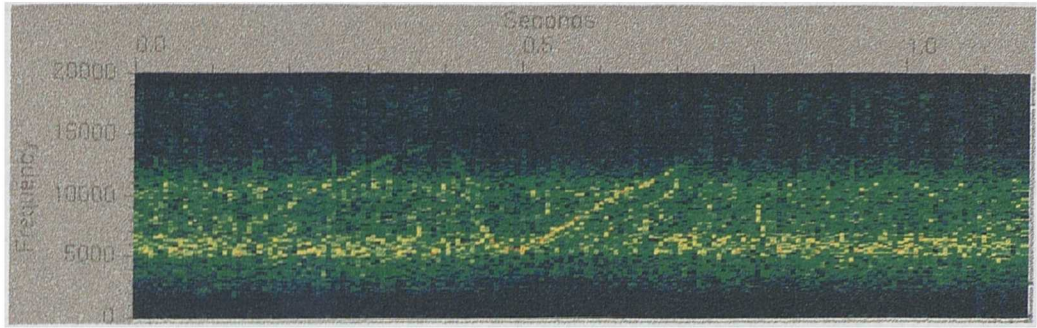
Figure 4-2: Whistle after partition equalisation

click train tended to be contained within one time partition on the spectrogram (5.8 ms at a sample rate of 44.1 kHz and a partition size of 256 samples, which was assessed to be good time partition for viewing the whistles). Since the spectrum of each click was roughly uniform over the frequency range of 0 to 22 kHz used in the spectrogram, it was decided to normalise the energy contained in each partition which should remove the echolocation clicks from the spectrogram.

The average intensity for each partition was calculated, and the ratio of this factor to the desired average energy was used to modify the data for all frequencies in that partition (see Figure 4-2). Thus if a time partition consists of N frequency bins, each with an energy intensity $i(n)$, the mean intensity $\bar{i}$ can be calculated for each partition. The intensity values can be modified so that the mean intensity of the modified partition will be some user-specified normalisation intensity value, here termed $\tilde{i}$. This can be achieved using Equation 4-1, where $i'(n)$ represents the modified intensity values for the partition.

$$i'(n) = \frac{\tilde{i}}{\bar{i}} i(n) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (4\text{-}1)$$

Visual inspection of the resulting spectrogram showed an increase in the contrast between the whistle and the rest of the spectrogram, although fragments of the clicks were still visible for some frequencies. If the modified FFT was deemed a sufficient improvement over the original, then it was written to disk. The detailed algorithm used for this process is covered in Sturtivant & Datta, 1995 (see Appendix D). This technique was assigned the name 'partition energy equalisation' to distinguish it from later techniques that were developed.

47

Enhancing the Whistle

In many cases the partition energy equalisation routine did not give ideal results, since the click spectrum was rarely uniform over the entire frequency range of the spectrogram. Click spectra could vary considerably between different signals, and frequently within a single click train, which was possibly the result of the dolphin changing its orientation with respect to the hydrophone whilst producing the clicks. Lower frequencies, being less directional, would be represented more strongly when the animal was pointing away from the hydrophone relative to the higher frequencies contained in the pulse.

The use of intensities from all frequencies to calculate the normalisation factor proved to be unacceptable in a number of cases, due to the non-uniformity of echolocation click spectra and other broadband noises. However, if a smaller frequency range were chosen, then the variation in intensity was much reduced. A technique was developed whereby the normalisation factor for each point intensity on the spectrogram was modified based on a calculation of the average intensity of a small range of frequency bins surrounding it. To clarify, for each time partition and for each frequency bin n, the mean of the intensities around n in the range $n \pm m$ is calculated, and this value used as the normalisation factor for that point. The main difference between this method and that of the previous one of partition energy equalisation, is that in this method the normalisation factor needs to be recalculated for each point rather than being applicable to the entire time partition. The equations for calculating the modified spectrogram now become:

$$\bar{i}_m(n) = \frac{1}{2m+1} \sum_{k=n-m}^{n+m} i(k) \dots\dots\dots\dots\dots\dots\dots\dots\dots(4\text{-}2)$$

$$i'_m(n) = \frac{\tilde{i}}{\bar{i}_m(n)} i(n) \dots\dots\dots\dots\dots\dots\dots\dots\dots(4\text{-}3)$$

The frequency bin range m that was used as the range over which to calculate the mean intensity for each point, was termed the mask and had a large effect on the appearance of the processed spectrogram. The key use of this mask led to the term 'masked energy equalisation' being used to describe this technique. For a mask size of 0, the entire spectrogram was flattened with all features disappearing, since each
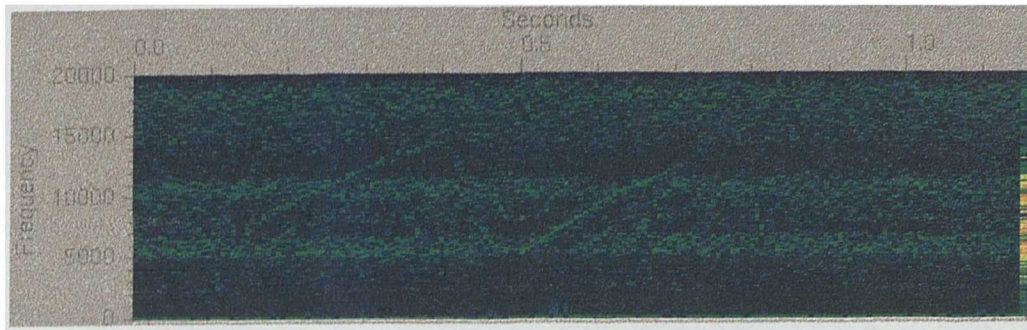
Figure 4-3: Whistle after mask equalisation

point in the spectrogram was reassigned the normalisation average value $\tilde{i}$. A mask with m equal to the number of frequency bins N in each partition, produced the same result as for the previous partition energy equalisation routine. Small values of m would remove the click, but if the mask was smaller than the range of frequencies contained in the whistle for one time partition, the whistle was also removed. Optimal results were obtained when m was small enough so that the click spectrum changed little over the mask area, but large enough that the whistles occupied a small range of frequencies for each time partition compared to the mask. Figure 4-3 shows a modified spectrogram with m set to 30, which was the default value chosen for this parameter. The mathematical background to this algorithm is given in Sturtivant & Datta, 1995 (Appendix D).

One effect of this routine was to create a 'quiet' zone in the frequencies surrounding a whistle for a distance of m frequency bins. This was due to the high intensity frequencies contained in the whistle increasing the average intensity calculated for the m surrounding elements. This effect proved useful when attempting to detect and isolate whistles.

Finding the Whistles

The masked energy equalisation routine modified the spectrogram so that any whistles that were present were much more apparent to visual inspection. However, the problem still remained to be solved of the detection of whistles that are scattered within long periods of background noise. The analogy with visually identifying areas of the spectrogram containing whistles provided a natural starting point for designing an algorithm to duplicate this task. The incoming signal from the recordings was

sampled either through a PC sound card at 44.1 kHz, and the resulting stream of data made available to the detection routine, or sampled from a different system and then transferred as a sound file.

An FFT of the sampled signal was calculated to convert the signal into the time-frequency-intensity domain. Although a time partition of 256 samples (corresponding to 5.8 ms at 44.1 kHz) produced a suitable time-to-frequency resolution for viewing the shape of an entire contour, for the detection routine a smaller partition size of 32 samples (corresponding to 0.73 ms) was used. This briefer partition time resulted in frequency bins with a width of 689 Hz, and thus whistle contours (which change only slowly in frequency) made the transition between adjacent bins only infrequently.

In order to reduce the contribution to the signal made by impulsive noises (such as echolocation clicks), the masked energy equalisation filtering technique was used to enhance signals with narrow frequency bandwidths. Even after filtering, noise from one time partition to the next was still present on the spectrogram due to components in the background which had a narrower spectrum than that of the echolocation clicks that the routine was designed to filter out. Taking two averages of the resulting filtered signal using an exponential decay rule (Equation 4-4) reduced these unwanted, short-term components of the signal. In this equation x is a series of values presented sequentially, x_{ave} is the current moving average of these values, and x'_{ave} the next value for the moving average after modification by x.

$$x'_{ave} = \alpha x_{ave} + (1-\alpha)x, \quad \text{where } 0 \le \alpha \le 1 \ \dots\dots\dots\dots\dots\dots(4\text{-}4)$$

It can be seen that this equation describes a signal following function that derives its next value to a greater or lesser extent from the previous value of the moving average x_{ave}. The constant α dictates how closely the function follows the original data — α close to 1 will produce a function that only slowly tends to react to changes in values, and α close to 0 a function that follows the original values very closely. If we consider a sudden change in the value of x from, for instance, a value of 1 to 0, then the value of the moving average at the next iteration will be α, and subsequently α^2, α^3, etc., producing an exponential decay tending to this new value with a specific 'half-life' (in this case log 2 / log α).

50

The long term background noise of the signal was found by using a large value for α that produced a 'half-life' of several seconds, and that for the averaged current signal by using smaller values for α giving a half-life of a few milliseconds. The instantaneous (filtered) signal spectrum $i_{inst}(n)$ could then be calculated as the difference between the averaged current signal and the long-term background noise. The mean, $\overline{i_{inst}}$, and standard deviation, σ_{inst}, of the instantaneous signal spectrum were calculated, and 'peaks' in the spectrum defined as frequencies whose intensities exceeded some threshold multiple m_{thresh} of standard deviations above the mean (Equation 4-5).

$$peak(n) = \begin{cases} 0, \text{ if } & i_{inst}(n) < \overline{i_{inst}} + m_{thresh}\sigma_{inst} \\ 1 & \text{otherwise} \end{cases} \quad \dots\dots\dots(4\text{-}5)$$

However, using $peak(n)$ alone to indicate the presence of a whistle was not particularly satisfactory since a high incidence of false alarms occurred, and increasing the value of α for the averaged current signal calculation reduced the algorithm's ability to detect rising or falling whistle contours. Rather than use this method on its own, an additional list $length(n)$ was kept of which frequency bins contained peak values from each time partition, and the number of time partitions for which the peak had been in existence at that or adjacent frequency bins (Equation 4-6). If no peak was indicated then $length(n)$ was reset to zero. Thus a potential whistle peak could be followed even if it changes frequency bins, and then flagged as a 'detected' whistle in $detect(n)$ when the peak had been in existence for a certain user-specified time $length_{thresh}$ (Equation 4-7). This method significantly reduced the number of false alarms without increasing the number of undetected whistles.

$$length'(n) = \begin{cases} 0, & \text{if } peak(n) = 0 \\ & \text{else} \\ 1 + \underset{n-1 \leq p \leq n+1}{\text{Max}} [length(p)] \end{cases} \quad \dots\dots\dots(4\text{-}6)$$

$$detect(n) = \begin{cases} 0, \text{ if } & length(n) < length_{thresh} \\ 1, & \text{otherwise} \end{cases} \quad \dots\dots\dots(4\text{-}7)$$

The *detect(n)* variable thus was set (i.e. a value of 1) when a whistle was present and contained in frequency bin *n*, and reset (i.e. a value of 0) when no whistle was detected in that frequency bin. Since the software was not fast enough to carry out whistle detection and store flagged data to disk, a log was made of those times when a whistle detection was made. The logged times indicated a duration starting slightly before the initial detection and ending after detection was lost, in order that none of the whistle contour was cut off. Just as a human eye requires several time partitions of a whistle to be displayed before a whistle's presence could be detected, so the software routine also required several sets of data before an increase in the intensity at a frequency could be identified as above normal background levels. A second pass of the program over the recording was subsequently required to sample the data onto the computer's hard disk at the times specified by the log.

The performance of the whistle detection routine proved to be very reliable. Out of a total of 52 whistles from common dolphins, which had been detected by ear, 50 of these were also found by the software routine when they were played back. Unfortunately, this algorithm also detects other tonal sounds. Sustained high repetition rate click trains, and even on one occasion and anchor chain clinking, have both been identified as potential whistles by the software. In generally, the excellent performance of this algorithm at whistle detection outweighed the cost of such false alarms.

Summary

The goal has been achieved of automatic detection of the presence of a whistle on a recording or from a live source. Although the whistle cannot directly be sampled to disk on computer as they occur, the times at which they occur on a tape can be logged, and then the tape replayed and those times at which whistles are present sampled for further analysis.

Two routines were developed to remove some of the sounds that contributed to the background noise in the recordings, based on the wide-band short duration nature of the echolocation clicks and several other noise sources. Normalisation of the total energy contained in each time partition of the spectrogram increased the visibility of the whistles, but often left some remnant of the echolocation click behind due to their

non-uniform spectra. The use of a narrower frequency range when calculating the normalising factor on a per element basis increased differentiation still further between whistle and background by surrounding the whistle with a quiet 'trough'.

Whistles were detected by using a short time partition and wide frequency bins on the spectrogram, which ensured the energy contained within each whistle rarely moved from one frequency bin to the next on consecutive time partitions. Long term and short term exponential running averages aided removal of both steady spectrum and short tonal background noise sources. Peaks in the resulting spectrum were traced, and if a contiguous chain could be found for more than a certain threshold duration, the presence of a whistle was indicated at that time. When the chain ceased, the whistle was indicated as having finished.

Once those portions of recordings that contain whistles have been isolated, and the signals sampled to computer disk, further and more detailed analysis can take place. The whistle contour is required, which consists of the sequence of time-frequency-intensity tuples that make up the whistle. However, the extraction of this contour is far more difficult than the detection of the whistle, and will be described in the next chapter.

References

Sturtivant, C. & Datta, S. (1995). "Techniques to isolate dolphin whistles and other tonal sounds from background noise." *Acoustics Letters* 18(10):189–193.

Chapter 5: Extracting the Whistle Contour

Introduction

The point had been reached where whistles contained on a recording could be located and sampled to a computer file. During the development of the necessary algorithms to detect a whistle's presence, filtering routines had been written that would both reduce the contributions made to the spectrogram by unwanted sounds and also enhance the whistle components. These filters additionally proved useful when the time-frequency-amplitude information for the whistle (hereafter referred to as the 'whistle contour' for brevity) was to be extracted from the signal.

Simplified, the changes of the whistle's fundamental frequency and amplitude with time were required from the spectrogram. Given a low noise level and a clear whistle from a single animal, the whistle's contour could be extracted by choosing the frequency bin with the highest energy at each time partition on the spectrogram. Of course, this 'perfect' recording very seldom occurred, since even when recordings were made in oceanariums, water noise, reverberation, and vocalisations from other dolphins were often present and would produce higher instantaneous intensities with frequencies other than those contained by the contour.

The three-dimensional viewing routine described previously again proved an excellent way of conceptualising the problem, and based on observations made of whistle spectra with this utility, three differing methods of finding the 'best route' that followed the whistle contour through the spectrogram were attempted.

The Shortest Path through a Whistle

By studying several spectrograms of dolphin whistles, it was deduced that for a high intensity whistle in lower intensity background noise, the chain of points consecutive in time through the spectrogram that maximised the sum of intensity values would contain the whistle contour. In a perfect case, one would then simply link up those points within each time partition to find the contour, but as already stated, this ideal was seldom encountered in unmodified spectrograms from practical signals. The

partition energy equalisation and masked energy equalisation filters described in the previous chapter improved the signal to background noise ratio, but stray noises would still often occur with a higher intensity than those found within the whistle contour even in these modified spectrograms. If some restrictions were placed on the chain of points, however, a workable solution to the whistle contour extraction problem could be found

Directed Graphs and Searching

When the spectrogram contains high intensity values that do not fall upon the whistle contour, one can still subjectively perceive which parts belong to the whistle and which to other noises. We often mentally discard points that cannot form part of the whistle, and if this can be incorporated into a whistle searching procedure, then the new algorithm should be more tolerant of noise. One such way might be to search through the spectrogram for the route that maximises certain desirable attributes.

The spectrogram can be described as a rectangular matrix of points, or nodes, consisting of the frequency bins for the time partitions, each of which has an associated value — its intensity. If we wish to find the globally optimum route through the nodes, we need the route through the spectrogram that maximises the sum of the intensities of each point. The route through the nodes has a few restrictions, in that consecutive nodes must appear in consecutive partitions. This restriction on the route can be conceptualised as allowed 'edges' for transitions between the nodes. In this case the edges are not bi-directional, since one can only move forward in time through the spectrogram.

Solving this initial problem is quite simple, and would involve merely finding the node with the maximum intensity in each partition. This solution, however, does not really model the way we subjectively trace out whistles, since it take no account of the separation in frequency of consecutive nodes. A value, or 'weight', could also be associated with each edge that could make small frequency transitions more favourable than larger ones. The aim with this model would be to maximise the sum of the weights of both nodes and edges traversed for a route from one side of the spectrogram to the other. This type of problem falls into a well-studied class of graph theory whereby an optimum route through a graph of weighted edges and nodes is

required (see e.g. Knuth, 1973). In this case, where the aim is to calculate the most profitable route through a weighted graph, it is known to be computationally expensive to solve in all but the simplest of cases. Unfortunately, in our problem the number of edges from one node to the next is equal to half the size of the FFT partition. Consequently, finding the optimum solution from a search of all possible routes would be unfeasible in terms of computing time for any but the shortest of signals.

There are, however, methods of restricting the number of edges between nodes that needed to be considered. One simple method is to restrict the frequency choice for a subsequent time partition, thereby reducing the number of edges from one node to others in the next partition and also the calculation times involved. The next contour point in the chain was thus chosen from a range (defined by the user) centred on the frequency value of the previous point. The algorithm's progress could be monitored on the screen by a set of light blue pixels superimposed over the spectrogram. The details of this algorithm can be found in Appendix B.5.

Discussion of the Shortest Path Method

This algorithm for tracing the whistle's contour was found to be unsatisfactory for a number of reasons. The main one of these, as for the minimum spanning tree approach to whistle representation, was that of computation time. Additionally, the algorithm would often fail to find any routes at all through the spectrogram, since the offset and threshold values did not allow the trace to reach the start of the whistle. In those cases where the whistle start was found, the algorithm would then spend long periods attempting possible routes to the right hand side of the spectrogram. Careful choice of offset and threshold values was needed to avoid computational effort approaching that of the unmodified algorithm. The high level of user intervention required hardly conformed to the goal of an automated recognition package.

The restriction placed on the choice of the next frequency point was rather artificial, and was chosen for convenience of programming. A distance measure based on frequency difference could be used in place of the allowable frequency range, which in combination with the intensity at that point could be used as the 'weight' for that particular node. This would allow for a more accurate following of whistle

contours when they changed rapidly in frequency. One can visualise the original range of allowable frequencies as a 'square' function multiplied by the partition's intensities to give the weighting, whereas any function based on the frequency change could also be applied, for example an inverse square function, or a Gaussian function.

The time taken by the algorithm to traverse the signal before the whistle began, and from the end of the whistle to the end of the spectrogram, formed a large part of the calculation time. A start and stop time partition for the whistle could be estimated by comparing the signal to noise ratio for frequency bins, in a similar way to that used for whistle detection (see pervious chapter). This would allow the threshold and offset values to be tailored more precisely to the whistle, since no initial or terminating quiet period would need to be traversed. Again, this would increase the amount of user interaction required for whistle contour extraction, which conflicts with the goal of automated recognition.

In summary, the use of the offset and threshold parameters was found to be too artificial and restrictive. Although further modification could be made to the algorithm to provide a faster result, it was concluded that the computational effort involved with this technique was too high, and less intensive and recursive methods would give results more quickly and with no less accuracy.

Ridge Following

The recursive 'shortest route' algorithm was not greatly affected in its accuracy by the use of the partition or masked energy equalisation background noise filters, although visually these routines seemed to have a significant effect on the signal to noise ratio of the spectrogram. Background noise in these processed spectrograms had been reduced by the removal of impulsive sounds such as echolocation clicks, and, with the masked routine, a 'quiet' area was left around any narrow band sounds such as whistles. Whereas the shortest path algorithm recursively searched a range of possible whistle contours through the spectrogram, the mask filtered spectrogram often contained a clearly delineated contour surrounded by areas where the intensity values had been suppressed by the higher intensities contained in the whistle contour. In this case the recursive search used by the shortest path algorithm could be replaced by a much simpler method of following the maximum intensity value once a point on the

whistle contour had been located. If one refers back to the three-dimensional viewing routine, this method is the equivalent of following the 'ridge' that forms the whistle contour after filtering.

An algorithm was developed that traced out the whistle for consecutive time partitions in the mask filtered spectrogram, resetting to zero all those frequency bins considered not to be part of the whistle. It again employed a user-definable restriction in frequency changes in its algorithm to extract whistles in a similar way to that of the shortest route method, and had an option to write to disk the spectrogram data, masked to those frequencies surrounding the whistle. Programming details of the algorithm can be found in Appendix B.6.

The start of the whistle was determined by searching for intensities higher than a value specified by the user. Although by no means foolproof, this simple technique often proved sufficient due to the effects of the mask filtering routine. If due to excessive noise a problem occurred in finding a suitable value to locate the start of the whistle, selecting that part of the spectrogram that contained only the whistle could often solve this. The end of the whistle was identified when the signal to noise ratio in the next frequency search range fell below a user defined value. When compared to the offset used with the exponential intensity average, this method was much more natural and it was easier for the user to estimate the value that should be specified.

Discussion of the Ridge Following Technique

The execution time of the ridge following routine was typically much less than one second, thus making it far more suitable for mass analysis than the shortest path routine. Although a simplification of the shortest path method, in many cases it was sufficiently accurate to extract the whistle's contour from the filtered spectrogram.

The maximum number of frequency bins that could be traversed between consecutive time partitions was set to a user definable limit, which again entailed user supervision of what should ideally have been a completely automatic process. In practice, a value of ±3 for this limit could be applied to the majority of whistle contours, and only those contours that had a very rapid frequency modulation required higher values. An alternative method to deal with these whistles was to decrease the

time partition of the FFT for the spectrogram, with the consequent loss of frequency resolution, which would allow the ±3 bin limit again to be sufficient.

One drawback to this routine was its simple approach, although ironically this was also one of its appeals, since relatively simple obstacles could degrade the algorithm's accuracy. If a break occurred in the contour, the algorithm would cease searching and only identify the first contiguous part of the whistle. Since the algorithm's signal to noise threshold for indicating whistle termination used an exponential averaging technique (described previously), it was sometimes possible to pick up the second whistle part by lowering this value and thus having the routine search further before indicating the end of the whistle. This solution, however, would often incorrectly indicate a path that deviated quickly from that expected of the whistle as the time between the two whistle parts increased, and could only rarely be used successfully.

Whistles from separate dolphins were often heard concurrently, and whistle contours were observed that crossed over each other in frequency. This caused the algorithm to follow whichever of the two contours contained the more intense values at the crossover point, regardless of whether the new direction deviated greatly from its previous course. In this situation, the best that this technique could offer was to identify the loudest of a number of whistles, and at worst it was incapable of correctly identifying any.

Inertial Following

Differentiating between two simultaneous whistles at the crossover point was one of the failings of the simple ridge following routine. Separating the contours seems simple using subjective visual analysis, since we understand that contours tend to continue in the same direction with few sudden turning points. One way of modelling this observation is to attempt to have the tracing algorithm resistant to changes in the 'direction' of the contour. For example, if the frequency of the contour is changing at a specific rate with time, then those subsequent frequency bins which maintain that rate will be viewed as more probable than those which cause the rate to change. A number of techniques were investigated, and the idea of assigning an 'inertia' to some

'whistle following point' seemed the most natural, and indeed seemed to give the best results (see also Sturtivant & Datta, 1995).

A routine already existed for detecting the presence of a whistle, and had been used previously in the whistle detection part of the software (see an earlier chapter). This routine was employed to find a point on the whistle's contour, although due to the exponential averaging that was employed to detect those relatively stationary frequencies contained in the whistle, the point of detection was often several time partitions after the beginning of the whistle itself. The contour thus had to be traced both forward and backward from this point.

The current (exponentially time averaged) direction of travel of the point following the contour was assigned an 'inertial' factor. High values of the inertial factor would strongly encourage the selection of further points on the contour which were consistent with the current direction of travel, whilst lower values would allow high intensities away from this path to be considered. A mathematical treatment of this method will be given as part of a more detailed explanation.

The point P_f (initialised to P_{init}, the point of detection) was used to track the current position in the spectrogram, the vector V_{xy} (initialised to zero) indicated the current direction of travel, and the inertial component $\alpha_{inertia}$ (set between 0.0 and 1.0) indicated the algorithm's sensitivity to sudden changes in direction. If the inertial factor were set to zero, the tracking used no inertial component, and the next point would be determined independently of the current direction of travel. A factor close to 1.0 would indicate that the next point should be chosen almost entirely based on which point continues the current direction of travel, irrespective of surrounding intensity values.

The contour is first traced backwards from the point of detection. New velocity values $V'_{xy}(n,t)$ are calculated for all points in the previous time partition within a specified number of frequency bins, in addition to points immediately above and below P_f according to Equation 5-1.
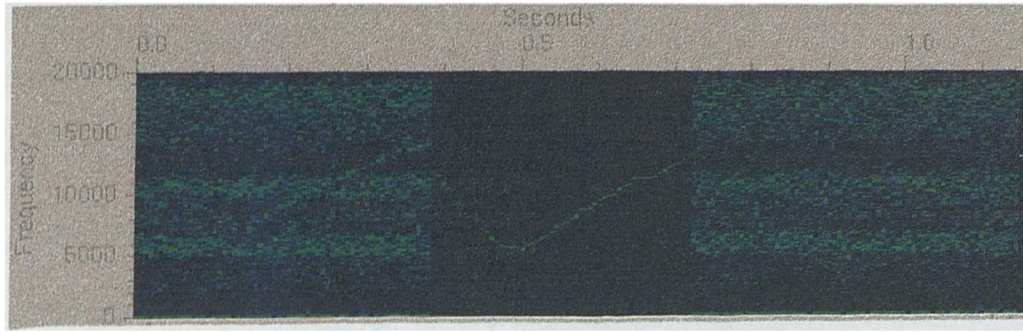
Figure 5-1: Mask filtered whistle with inertial following of the main contour

$$V'_{xy}(n,t) = \{\alpha_{inertia}V_{xy}\} + \left[(1-\alpha_{inertia})\frac{\Delta}{|\Delta|}i(n,t)\right] \dots\dots\dots\dots(5\text{-}1)$$

where Δ is the direction vector from P_f to $P(n, t)$.

The equation is split up into two components — the inertial part (given in curly brackets), and the intensity driven part (given in square brackets). The inertial part of the equation simply continues the current direction of travel, but the intensity driven part requires further explanation. The intensity $i(n,t)$ at each point can be visualised as the force that pulls the direction of travel away from its current course. The intensity was multiplied with the unit vector from the current point to that new point ($\Delta/|\Delta|$ in the equation) to give a direction to the intensity, and the direction of travel modified by the inertial factor. Thus, the magnitude of the current direction of travel V_{xy} is dependent on the intensity of the whistle contour that it is following — the more intense the whistle contour, the higher the vector magnitude. When all calculations for $V'_{xy}(n,t)$ have been made (for those points above and below the current point, and within a user defined range in the next partition), the point with the largest magnitude is chosen to be the next point on the contour. This series of events continues until the signal to noise ratio of the frequency bins surrounding the current point falls below a pre-set threshold. The algorithm is then used again to trace the contour forward from the detection point. It should be noted that this algorithm does not choose the point to which the vector points, but to the point that gives the largest vector magnitude. In this way the vector will trail the actual direction of travel.

Figure 5-1 shows the result of using the inertial following algorithm on the whistle shown in the previous chapter. Where the frequencies of the whistle have

61
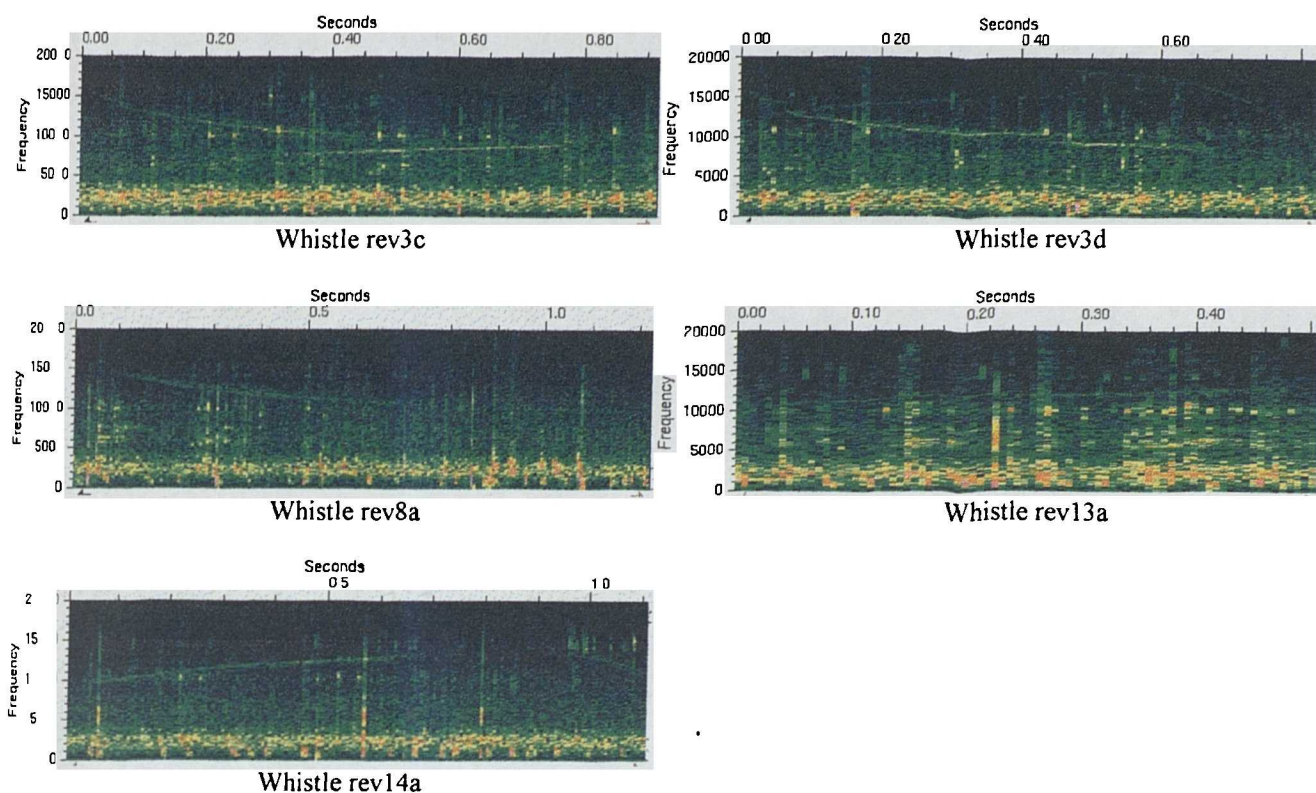
Figure 5-2: Five whistles that could not be satisfactorily extract by inertial following.

been isolated, the intensities for all other frequency bins in that partition have been reset so that only the contour remains (the area surrounded by black). No start point for tracing was found on the first part of the whistle, due to a combination of low signal to noise and high rate of change of frequency.

This method has a number of advantages for contour tracing that the simpler methods were unable to emulate. If the intensity of the contour falls, the magnitude of the direction of travel vector will also fall, allowing smaller intensity values to alter more easily the track that is followed. For a strong, intense contour, points away from the current direction of travel will need a correspondingly higher intensity to pull the track in that direction. The degree to which the track is smoothed out is dependent on the inertial component of the equation, with the effect that noise close to the whistle contour might not pull the track away from the true line to any large extent. After the end of a contour has been determined, the search for a further initial point can be made since a break in the whistle contour was sometimes found to occur.

Inertial Following Performance

The performance of the inertial following algorithm was assessed by attempting to extract the contours of 17 whistle trains from common dolphins, which contained a total of 52 separate whistles that could be visually distinguished. Of these whistles, a total of 47 were satisfactorily extracted, which was defined as deviations from only minor portions of the contour, with no alteration in the list of segment types.

Of the five whistle contours which could not be accurately extracted (shown in Figure 5-2), four were due to a poor signal to noise ratio, and the remaining one was due to the algorithm not detecting the last part of a whistle after a short gap. The first whistle in 'rev3c' contained a weak component. The initial detection point algorithm could not find this part, since the contour was not loud enough to be detected when changing so quickly in frequency. Only the second whistle in the signal was extracted. Whistle rev3d contained a weak simultaneous whistle at a higher frequency than the more dominant whistle, and only the lower one could be extracted. The whistles in rev8a and rev13c likewise contained whistles with poor signal to noise ratios. The last part of whistle rev14a was not detected after the break in the contour, probably due again to too low a signal to noise ratio for the frequency change.

Discussion of the Inertial Following Technique

Although the user was still required to make some indications about the starting point signal to noise value, number of frequency bins for the search, and inertial factor, these values could be left unchanged during the majority of contour extractions. From this point of view it is far more automatic than the previous two methods.

The inertial technique is fast, completing extraction of the contour of a typical whistle of a one second duration in less than half a second. No backtracking or recursion routines are contained in the algorithm to affect its performance, and it simply carries out a search over a limited part of the spectrogram ahead of (or behind) the current tracing point.

Concurrent whistle contours that cross each other in frequency can be separated, since the algorithm will attempt to continue in its current direction of travel at the cross-over points unless the second, unfollowed, whistle has a much greater intensity. Where this intensity difference causes a problem, it would be a simple task

to construct a routine which would flatten to background noise all those intensities on points on a predetermined contour through a spectrogram, thus allowing a less intense contour to be subsequently to be extracted. Currently, removal of a contour in this manner has not been attempted.

Another possible method of solving the overlapping contour problem might be to add the intensity of the contour point to the inertial equation. This would produce a three-dimensional direction vector, in which sudden changes in the intensity of a contour would be less desirable than a smooth continuation of the current time-averaged intensity changes. In general, the intensities contained in the mask filtered spectrogram will be different from those contained in the original signal. In this case, the filtered intensity values would be used as the multiplier for the non-inertial part, and the original intensity values for the direction vector, as follows:

$$V'_{xyi}(n,t-1) = \left\{ \alpha_{inertia} V_{xyi} \right\} + \left[(1-\alpha_{inertia}) \frac{\Delta_{xyi}}{\left| \Delta_{xyi} \right|} i'(n,t-1) \right] \qquad (5\text{-}2)$$

where $\Delta_{xyi} = (P'_{f_x} - P_{f_x}, P'_{f_y} - P_{f_y}, i(P'_f) - i(P_f))$, i represents the original signal intensity values, and i' the filtered intensities.

In summary, the inertial technique mirrored many aspects of visual contour tracing, with the exception of the matching of changes in intensity across cross-over points, and also the ability to mentally remove contours already traced from consideration of further contours. Both of these techniques could be applied by the techniques outlined above.

Unmasking the Whistle

Once a whistle's contour had been determined, the intensity values it contained usually had been filtered and so the intensities from the original signal could no longer be determined. However, the points on the contour could be used as a mask that could be applied to the original spectrogram data if this had previously been saved to disk. This 'unmasking' process allowed the whistle's precise frequency-time-intensity contour to be passed on for further analysis (Figure 5-2).

Signal Reconstruction

It would be reassuring to confirm that the whistle contour extraction process had produced the results that were expected, and one way of achieving this would be to play back the extracted contours though a loudspeaker. The isolated whistles could then be compared by ear with the original recording to ensure that the process had been carried out correctly and that no portions of the whistle had been neglected or extraneous noise misidentified.
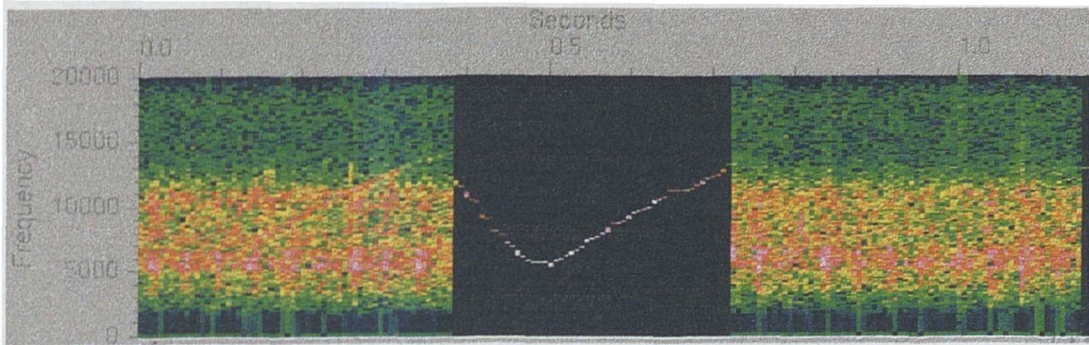
Figure 5-2: Extract whistle contour used as a mask on the original data.

The process of reconstructing a signal from the spectrogram data can be achieved by calculating the inverse Fourier transform of each time partition, and subsequently concatenating the resultant signals. The inverse Fourier transform can be determined by the following relationship:

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{2\pi ft}dt$$

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{-2\pi ft}df \quad \text{..} (5\text{-}3)$$

where $H(f)$ is the Fourier transform of the signal $h(t)$. In discrete form, where $H(f_n) \approx \Delta H_n$, for sample interval Δ,

$$h_k = \frac{1}{N}\sum_{n=0}^{N-1} H_n e^{-2\pi ikn/N} \quad \text{..} (5\text{-}4)$$

Comparison of this equation with that used for the fast Fourier transform (see equation 3-1, chapter 3) indicates that a similar method can be used to calculate the inverse FFT, if a scaling factor of the reciprocal of the partition size is included. The FFT routine was modified to calculate either the normal FFT of a signal, or to reconstruct the signal from the spectrogram data. Since the spectrogram intensities

are stored as power values, the complex data indicating phase and amplitude are required from the original FFT for those frequencies that are to be included in the final signal. If the power spectrum is used to reconstruct frequency information simply with real components, the resultant signal will have none of the time varying features contained within each time partition of the original. Additionally, the signal may make a sudden transition across a time partition boundary, causing a noticeable broadband 'click' when played back.

These effects could be minimised by storing both real and imaginary components of the FFT throughout each modification made to the spectrogram, and using these values when reconstruction of a signal was desired. Since one half of the FFT of a wholly real signal is the complex conjugate of the other, necessitating storage of only one half of the transform, the other half can be reconstructed when the inverse transform needs to be applied.

When replayed, the modified signal often would still contain some regular interference. Investigation indicated small discontinuities in the signal at time partition boundaries. This was to be expected, since many of the frequencies contained in the original signal had been discarded in the final contour, producing changes in the signal amplitude throughout the partition, including the start and end. The use of a 'window' to mix consecutive partitions (e.g. overlapping triangular windows, Gaussian windows, etc.) would remove these transients if playback was necessary of a signal that changed smoothly in frequency.

Volumes and Data Formats

The partition and masked energy equalisation filters normalised the frequency intensity values contained in a signal, often reducing the overall volume by several tens of decibels. If such a filtered signal were replayed, it was often difficult to perceive the signal since it was so quiet.

Another problem occurred with the format of computer sampled sound files. The sound format required by the software was a stream of signed 16-bit integers, however not all signal capture software produced this type of file. There were

additional problems encountered with sounds recorded on computers other than IBM PC's.

Two utilities were written that affected the signal (as opposed to the spectrogram) to overcome these problems. The first was developed to overcome problems of different sound file formats, and the second to adjust the volume of a signal so that it could be heard more easily.

Converting Sound Files for Analysis

There are several different formats available for sound files, many of which contain further sub-formats to take advantage of advances in sound cards. Some of the more common ones include '.wav' or RIFF, '.au' files typically produced from Sun workstations, '.voc', '.raw', '.snd', and '.aiff'. Fortunately, there is available an increasing number of software packages which can convert between many of these diverse formats. The one used as the standard in this project was the '.raw' format, since it provided a simple signed 16-bit integer for each sample, and was one of the formats available on many of the conversion routines.

A problem often occurred when sounds were brought from non-IBM PC systems for analysis. The sound format was stated as signed 16-bit integers, but when analysed they would seem to contain random noise with no discernible tonal components. The problem was traced to the method by which computers store multiple-byte numbers, in that some store the most significant byte (i.e. the one which has the largest effect on the integer's value) before the least significant byte, and others vice versa. This was solved by writing a simple program to exchange bytes within each 16-bit number throughout the whole of a sound file.

Those formats that stored the number as an unsigned integer also had to be accommodated. In this case, changing the highest bit (sometimes called the 'sign' bit) had the effect of moving the values into the appropriate range. The highest unsigned value (all bits set to 1) would be moved to the highest signed value (sign bit 0, but all other bits set 1). The middle unsigned value (most significant bit set 1, all others 0) would become the signed value zero (all bits 0), and the unsigned value of 0, the lowest (all bits zero), would become the smallest signed number (sign bit 1, all others zero).

Normalising a Sound's Intensity

Signals reconstructed from filtered spectrogram data were often so quiet that they were difficult to distinguish when replayed through speakers connected to the computer. Comparing the final signal with the initial signal was found to be a good method of validating that the routines had indeed extracted or enhanced the whistle correctly. In order to achieve this, some method was required that restored the final signal either to a similar volume to the original recording, or to some other normalised volume setting.

A scaling routine was written which operated by searching through the 16-bit signal file to determine the maximum absolute value used. This maximum value was used to calculate a factor by which all of the data was scaled. The scaling factor S_f was calculated as follows:

$$s_{max} = \max[|s_i|, \forall i]$$
$$S_f = \frac{2^{15} - 1}{s_{max}} \qquad\qquad (5\text{-}5)$$

where s are the discrete values for the signal data. All values within the signal are subsequently multiplied by the scaling factor S_f. A 16-bit signed integer can take values from -2^{15} to $2^{15} - 1$, thus the scaling factor will never produce a value outside this range.

Summary

Three techniques were developed for the extraction of the whistle contour from a spectrogram. Filtering routines, which have been previously described, were used to aid the routines in the identification of the whistle contour, although certain of the routines profited more from these than did others. An extracted whistle contour could have the intensities contained in its frequencies restored from the original unfiltered spectrogram, thus retaining the complete time-frequency-intensity contour, which was the primary aim of this task. Such contours could then be stored on disk for use in later analysis.

The 'shortest path' routine used a variation on a standard graph searching algorithm, in which various routes through the spectrogram are recursively calculated.

The route with the best sum of weighted 'costs' is then indicated as the route that contains the whistle contour. Restrictions had to be imposed on the allowed frequency change from one time partition to the next in order to eliminate short, intense, narrow band noise being mistakenly identified as belonging to the whistle. The large computational effort that was required to process each spectrogram, and the problem of traversing the regions before the start and after the end of the whistle, restricted the suitability of this technique.

A simpler method was entitled the 'ridge following' routine, which took advantage of the lower intensity noise 'trough' that surrounded the whistle after the masked energy equalisation filter had been applied. Since this routine was non-recursive, its execution time was notably less than that for the shortest path routine. The user was often required to tailor the routine's parameters to extract each whistle most successfully, which was undesirable since one objective was to produce an automatic extraction process. Problems with the routine also occurred if there were short breaks in the whistle's contour, or if more than one whistle occurred simultaneously. These simultaneous whistles could either restrict the routine to only identifying one whistle, or in the worst case neither if the two contours crossed in frequency at some instant.

A further refinement of the whistle extraction method incorporated an inertial component in the whistle following routine. This caused the algorithm to continue to follow a contour in the direction in which it was previously heading unless larger intensities caused its path to deviate. The more sudden the change in direction, the higher the intensity required in order to cause the routine to take that path. Breaks in the whistle contour were also accommodated, since the routine would continue to search for another starting point after it had detected the end of a previous whistle.

These improvements allowed the problem of simultaneous whistles to be solved, since the routine would follow a contour correctly if it crossed that of a different whistle. Although more parameters were required for this inertial routine, in practice the user could often leave them at settings that would enable a large proportion of the whistles to be extracted. Contours that featured sudden changes in direction could quickly be extracted by lowering the inertial factor.

The whistle contour could now be extracted and isolated from other sounds contained in the original signal. The next task was to compare one contour with another, and to calculate a quantitative value for their similarity, which is covered in the next chapter.

References

Knuth, D.E. (1973) "The art of computer programming: Volume 3. Sorting and searching", published by Addison-Wesley.

Sturtivant, C.R., and Datta, S. (1995) "The isolation from background noise and characterisation of bottlenose dolphin (*Tursiops truncatus*) whistles", *J.Acoust.Soc.India* 23(4):199–204.

Chapter 6: Encoding the Whistle

Introduction

Spectrograms from which whistle contours could most easily be isolated typically contained time partitions of 256 samples, and whistles with duration frequently in excess of one second. If we consider a sample rate of 44.1 kHz, giving a time partition of 5.80 ms, a whistle with a one second duration will contain approximately 172 partitions. The comparison of two whistles of this length consequently becomes comparison of two sequences of 172 values, which, if timing differences between the whistles are to be taken into account, becomes a non-trivial task. Comparison with a small database of one hundred whistles becomes a task that might easily require tens of minutes of computation on current computers.

Clearly, it is the high data content in each contour that leads to such a high computational requirement. If we consider the number of required comparisons to be of the order N^2, a reduction in the data would have a large effect on the comparison time. The critical question is what method should be employed to reduce that data size that least affects the information content? This leads naturally to the question of what is the information content of each whistle? Since dolphins have been shown to be able to identify conspecifics (others of the same species) from recordings of whistles (Caldwell *et al.*, 1970; Caldwell *et al.*, 1971), information on identity must be present in these whistles.

Encoding Methods

Santhi Mahadevan (1994), and Brenda McCowan (McCowan, 1995; McCowan & Reiss, 1995) have recently presented two methods of whistle contour comparison. Mahadevan's study analysed recordings made of four captive bottlenose dolphins during free swimming echolocation trials, and compared five features of their whistle contours:

> duration;
>
> frequency (including start, end, lowest, and highest frequencies);
>
> number of frequency reversals;

slopes;

amplitude (including start, end, lowest, and highest amplitudes).

She found that the number of frequency reversals could be used to identify two of the dolphins, and that the duration of the whistle could be used to distinguish between the remaining two dolphins. The accuracy of identification was improved by including other features into the comparison, but these two were the most important in this study.

These techniques were applied to over a hundred whistle contours extracted from the Moray Firth data, and the results discussed at the 1995 European Association for Aquatic Mammals conference in Nürnberg, Germany (Sturtivant *et al.*, 1995). Initial cluster analysis comparing frequency reversals and duration was inconclusive, and no strong clusters were evidenced. Comparison of additional feature pairs also failed to indicate any strong clusters. The lack of *a priori* information about the distribution of features contained in the dolphins' whistles (which had been present in the captive study) severely hampered identification of individuals. Cluster analysis of more than two features at a time might lead to the formation of identifiable discrete clusters, although this was thought to be unlikely.

Brenda McCowan's technique measured the frequency at 20 equidistant points along each contour in order to encode the whistle. It should be pointed out that Mahadevan had also used a similar technique, measuring the slope of the whistle at 10 points, and had some success in identification from this information. McCowan constructed a correlation matrix using the 20 frequency points as variables, which resulted in similarity indices between whistles. She also used principal component analysis to reduce the number of variables that required consideration in the correlation matrix, and then K-means cluster analysis was applied to group the whistles into a minimum number of non-overlapping classes. Discriminant analysis was conducted on the resulting classes by removing up to 20% of the whistles from the classification calculation, and then reclassifying these whistles according to the new classification function. The χ^2 statistic for the number of correct classifications (relative to the original, all-data, classification) was used to indicate the success of the technique.

McCowan's clustering algorithms and statistical treatment of the data is quite rigorous, but relies on the 20 frequency points to contain all of the information on identity in the original whistle. Two of the stated benefits of this method are its insensitivity to relative durations and frequency shifts, since dolphin whistles can vary in these two parameters. However, it is conceivable that two whistles from different dolphins may have very different durations or frequencies, but that when these 20 points are 'normalised' they might have similar values. Such whistles might then mistakenly be classified together, whereas it would seem more prudent initially to classify them separately and then optionally merge the classes after further analysis. Once information on duration and absolute frequency has been discarded, it cannot later easily be retrieved. A case in point was Santhi Mahadevan's use of whistle duration in her study to distinguish between two dolphins, whose whistle contours appeared otherwise quite similar.

Whistle contours from the same dolphin can differ in duration and frequency, and this feature should be accommodated in whatever encoding algorithm is employed, but repeated whistles also need inclusion in the encoding. Our observations suggest that dolphins, especially when excited, seem to repeat a basic whistle continuously for a number of repetitions. Under McCowan's method, a whistle with two repetitions would have a very different observation vector (the 20 observation points) from a whistle with one, three or more repetitions, which might subsequently lead to them being classified separately.

Having studied the representational systems used by both McCowan and Mahadevan, an encoding system was developed that preserved the shape of a whistle's contour whilst at the same time reducing the data requirement to an acceptable level for comparison with a large database of contours. It should be noted that both Mahadevan (Mahadevan, 1994) and McCowan (McCowan, 1995; McCowan & Reiss, 1995) were analysing data in which the identification of the vocalising animal was known, whereas the Moray Firth data has no *a priori* identity information.

Encoding Based on Acoustic 'Shape'

The two methods described above made use of the shape of the whistle contour —
Mahadevan calculated features that were dependent on the shape, and McCowan by

Segment	Source
" U "	Moray 92 Tape 8 w 24b
" D "	Moray 92 Tape 7 w 4
" "	Moray 92 Tape 5 w 21b
" "	Moray 92 Tape 7 w 20
" "	Moray 92 Tape 7 w 36
" "	Moray 92 Tape 5 w 19a
" "	Moray 92 Tape 5 w 28b
" "	Moray 92 Tape 5 w 28c
" "	Moray 92 Tape 5 w 24
" "	Moray 92 Tape 5 w 23
" U "	Moray 92 Tape 5 w 26
" "	Moray 92 Tape 5 w 19b
" D "	Moray 92 Tape 5 w 25
" U"	Moray 92 Tape 5 w 28a
" U "	Moray 92 Tape 5 w 19c
"	Moray 92 Tape 5 w 18b
" '	Moray 92 Tape 5 w 18c
" "	Moray 92 Tape 5 w 21a
" "	Moray 92 Tape 5 w 28f
" "	Moray 92 Tape 5 w 31
"	Moray 92 Tape 7 w 16
" '	Moray 92 Tape 7 w 38
"	Moray 92 Tape 7 w 40
'	Moray 92 Tape 7 w 41
"	Moray 92 Tape 7 w 46
' '	Moray 92 Tape 7 w 47a
	M ray 92 Tape 7 w 48b
'	Moray 92 Tape 8 w 1
" '	Moray 92 Tape 8 w 16
'	M ray 92 Tape 8 w 19
' '	M ray 92 Tape 8 w 5a
D	M ray 92 Tape 7 w 50a
D	M ray 92 Tape 7 w 47b
D'	M ray 92 Tape 7 w 59
D	M ray 92 Tape 7 w 2
U D '	M ray 92 Tape 7 w 57
U	M ray 92 Tape 5 w 28e
U	M ray 92 Tape 5 w 36
U	M ray 92 Tape 7 w 23
U	M ray 92 Tape 8 w 2
U	M ray 92 Tape 8 w 24a
U	M ray 92 Tape 8 w 3
U U'	M ray 92 Tape 5 w 27
UDU'	M ray 92 Tape 7 w 22c
D U DU U DU "	M ray 92 Tape 8 w 8
D U	M ray 92 Tape 5 w 30
D	M ray 92 Tape 5 w 34
D	M ray 92 Tape 7 w 1
D	M ray 92 Tape 7 w 74
D D	M ray 92 Tape 8 w 20
D DU '	M ray 92 Tape 8 w 23c
D U D	M ray 92 Tape 8 w 11

Segment	Source
"DU D "	Moray 92 Tape 7 w 9
"DU "	Moray 92 Tape 7 w 48a
"DU- "	Moray 92 Tape 7 w 55
"DU-"	Moray 92 Tape 5 w 35
"U "	Moray 92 Tape 7 w 50b
"U"	Moray 92 Tape 5 w 18a
"U"	Moray 92 Tape 5 w 28d
"U"	Moray 92 Tape 8 w 15a
"U"	Moray 92 Tape 8 w 15c
"U "	Moray 92 Tape 5 w 44
"U "	Moray 92 Tape 7 w 45
"U "	Moray 92 Tape 7 w 73
"U- "	Moray 92 Tape 8 w 23d
"U "	Moray 92 Tape 8 w 6
"U- "	Moray 92 Tape 7 w 22b
"U- - - - - "	Moray 92 Tape 5 w 20b
"U "	Moray 92 Tape 5 w 20a
"U- "	Moray 92 Tape 7 w 69d
"U- D "	Moray 92 Tape 7 w 69a
"U- D"	Moray 92 Tape 5 w 43
"U- D"	Moray 92 Tape 7 w 67a
"U- D"	Moray 92 Tape 7 w 69b
"U- U-"	Moray 92 Tape 8 w 23a
"U "	Moray 92 Tape 5 w 29
"U "	Moray 92 Tape 5 w 45
"U-"	Moray 92 Tape 7 w 29
"U-"	Moray 92 Tape 7 w 37a
"U-"	Moray 92 Tape 7 w 37b
"U-'	Moray 92 Tape 8 w 23b
"U D"	Moray 92 Tape 7 w 54
"U D"	Moray 92 Tape 7 w 67b
"U D"	Moray 92 Tape 7 w 69c
'U-D- - "	Moray 92 Tape 5 w 15
"U-D- -"	Moray 92 Tape 5 w 10a
"U D- -"	Moray 92 Tape 5 w 13a
"U D- -"	Moray 92 Tape 5 w 13b
"U D -"	Moray 92 Tape 5 w 14
"U-D- -"	Moray 92 Tape 5 w 16b
"U-D- -"	Moray 92 Tape 5 w 17
"U D "	Moray 92 Tape 7 w 61
'U-D- -U"	Moray 92 Tape 5 w 12
"U-D -U"	Moray 92 Tape 5 w 16a
"U-D- D-"	Moray 92 Tape 5 w 11a
"U-D-D"	Moray 92 Tape 5 w 10b
"U D-U -"	Moray 92 Tape 5 w 11b
"U-DU-D"	Moray 92 Tape 8 w 5b
"U-U"	Moray 92 Tape 8 w 15b
'U-U-D"	Moray 92 Tape 8 w 10
"U-UD"	Moray 92 Tape 7 w 22a

Table 6-1: 'Alphabetical' ordering by segment type of the Moray Firth whistles. Segment key: '-' flat; 'D' down; 'U' up; and ' ' for a gap.

sampling at time-independent points on the contour. Since neither method maintains all features, information on the whistle's shape has necessarily been discarded, some of which might be important for use by one dolphin to identify another.

The contour encoding method here developed represented each whistle at two levels of detail. At the lowest detail level, each contour was split into a series of segments which indicated whether the contour's frequency rose, fell, remained approximately the same, or was absent for a short period of time. In this manner,

quite complex whistles consisting of many repetitions of one or a number of basic whistles could be represented by a small number of segments, each labelled as one of four types.

During trials of segmenting a number of whistles recorded from wild and oceanarium animals, the number of segments seldom exceeded ten, and for the majority of whistles this remained around the four to six range. McCowan's criticism that '… most techniques do not control for the frequency/time expansions and compressions found in the whistle contours of some species, such as the bottlenose dolphin …' (McCowan, 1995) does not apply here, since the same whistle shape will be represented by the same sequence of segments whether its duration is one second or one quarter of a second. The information stored for a whistle will depend upon the contour's complexity, and no compromises need to be made concerning the best number of sample points to represent all whistles.

The comparison of contours is thus translated into the comparison of short sequences of segment types, clearly a much simpler task. Simply placing these whistles into numerical order based on the their segment types is a useful aid in whistle classification. A study of 101 whistle contours from Moray Firth recordings (Sturtivant & Datta, 1996a and 1996b) indicated that such an ordering grouped similar whistles together, and one particular section from this list was highlighted. This section contained whistles of the 'rise, flat, fall, flat, break, flat' variety, which were quite similar when viewed. The data from this study has been reproduced in Table 6-1. When the contours had been extracted, it was noticed that two contours belonging to this type occurred simultaneously. Close examination indicated differences in the contours indicative of two dolphins, and it was suggested that this might be an instance of whistle mimicry in a wild dolphin population. Since this encoding technique identified shape on only a basic level, the small differences observed were lost, and thus classification by this method was incapable of separating them (which was perhaps the intention of the mimicking dolphin towards other dolphins).

More detail on the shape of the whistles was needed, since no indications of relative durations of segments or their internal structure were present in the simple label of, for instance, 'rising'. A second level of detail was added, allowing time duration and the frequency-time shape of the curve to be stored within each segment.

Thus the chain of segments could theoretically be used to reconstruct the original whistle, albeit without intensity information.

Segmenting the Contour

The determination of the 'segment type' at each point in the whistle contour required a moving average to be kept, as the discrete nature of the frequency bins would otherwise cause too frequent a movement between the 'rise', 'fall', and 'flat' states. The 'inertial vector' from equation 5-1 was used to indicate the smoothed direction of travel, although this introduced a 'lag' due to these same smoothing effects. An 'amplitude threshold' value was also required for determining when a contour was considered to be 'broken' or 'starting', which also had a large effect on the initial number of segments in to which the contour was divided. Despite fine tuning of these parameters both globally and on a whistle by whistle basis, the software often indicated short segments due to noise in the original signal which did not fit in to the overall shape of the contour. A typical whistle contour split up in to segments is shown in Figure 6-1 after initial segmentation.
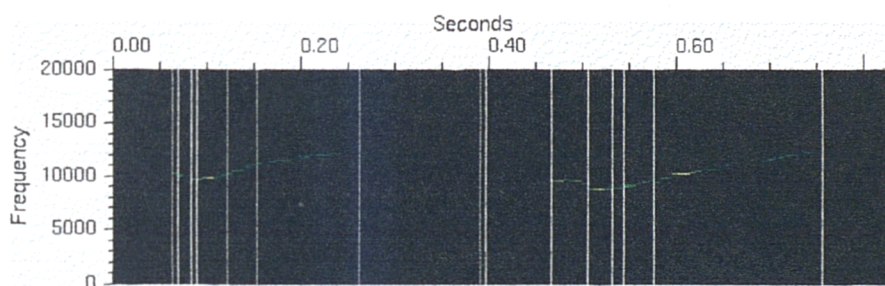
Figure 6-1. A typical whistle contour after initial segmentation

Combination of segments reduced this large number, where small segments would be merged with larger ones. In order that the least significant segments were removed first, the segment list was scanned to find the briefest segment, and this was merged into its two neighbours. If the two neighbours of the original (small) segment had the same segment type (as often occurred if the rogue segment was caused by noise), then they were also merged together. This operation was repeated until the smallest segment was larger than some threshold value, chosen to be 1 20th of a second. The 'tidied' segments from the Figure 6-1 example are shown in Figure 6-2.
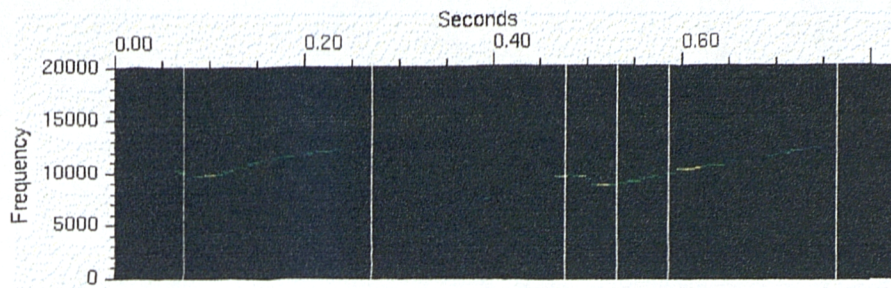
Figure 6-2: The same whistle after noise segments have been removed.

A problem with the segmentation of the contours can be seen due to deciding on the segment boundaries solely by a left-to-right scan of the contour. The segment boundaries are thus moved further to the right than would be expected due to the moving average technique used to determine the current slope of the contour (mentioned previously). This effect could be avoided by comparing these segment boundaries with those found using a right-to-left scan, and placing the boundaries at a central point.

Matching the Curve

Observation of the 'shape' of the whistle suggested that a compact representation could be employed to describe its salient characteristics. An algorithm was adopted whereby the whistle was split into segments, indicating whether the whistle was 'rising', 'flat', or 'falling' in frequency with time, or 'blank' indicating a break in the contour. Although suitable for quickly identifying which whistles were largely similar to each other, it was insufficiently accurate to distinguish, for instance, whistle mimicry. For each segment, its duration was recorded, and the data points contained in each were further specified using a quadratic equation to map the contour's shape and allow for more detailed comparisons. The quadratic equation employed was expressed in the familiar form shown in Equation 6-1.

$$f(t) = a_0 + a_1 t + a_2 t^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (6\text{-}1)$$

The values for t were normalised insofar as $t = 0$ at the start of each segment. It would also have been possible to normalise t over the duration of each segment, so that t would range from zero to one. Since the segment duration is also recorded, the original curve could be restored by using factoring t with this value, and subsequent comparisons of the a_1 and a_2 attributes would allow duration-independent

comparisons of segments for slope and rate of change of slope. However, since time normalisation can be calculated at a later stage, it was decided to store each segment with t varying as the partition number, which allowed simpler segment comparisons for actual time (as opposed to relative time).

In general, the quadratic equation will not exactly match all points on the extracted whistle contour, so some 'best fit' was required. Several methods for calculating the difference between curves are available, but the two that seemed most appropriate were the sum of the modulus of the differences, and the sum of the square of the differences. It was deemed important that the fitted curve should deviate little from the contour in its entirety. Any curve that lies far from the contour for a short length but matches well along the majority is less desirable than one where the curve lies less close to the contour but which doesn't have any outlying sections. With this criterion in mind, a least-squares fitting routine was decided upon.

The suggested method for calculating a least-squares match to a large set of data is by the use of singular value decomposition (SVD), and a brief overview of the technique's application to the problem will be presented here. A more detailed discussion is presented in Sturtivant & Datta, 1995 (Appendix E). A matrix containing the time information for each point was constructed according to the equation shown below (Equation 6-2). This matrix, termed the 'design matrix' **A**, has a very simple form when the time values are increasing regularly as in this case, but it is capable also of taking a sequence of values randomly scattered in time.

$$A_{ij} = \frac{(t_i)^j}{\sigma_i} \quad \text{..} \quad (6\text{-}2)$$

In this equation, i ranges over the number of time partitions in the segment, and j from 0 to 2 (i.e. for the three powers of t used in Equation 6-1, including t^0). Since the standard deviations σ_i for each frequency point were presumed to be the same, these were all set to 1.0. The standard deviation in this equation represents the weight of each frequency-time point in the construction of the curve, thus setting all of these values to unity (as a convenient value) will assign the same weight to all points.

Frequency information was placed in a column matrix **b**, points being entered in the same order as that which was used for the design matrix **A**.

$$b_i = \frac{f_i}{\sigma_i} \quad\text{...} (6\text{-}3)$$

Again, each σ_i was set to 1.0 to assign equal weightings to the points for the curve.

A precise solution to this problem might then be to find the values of **a** (which is the column matrix of the parameters a_i from Equation 6-1) which fit the relationship shown in Equation 6-4.

$$\mathbf{A} \cdot \mathbf{a} = \mathbf{b} \quad\text{..} (6\text{-}4)$$

However, since we have restricted to three (for the quadratic equation) the number of parameters that can be used to form the curve, the equation would have no precise solutions in this case. The 'fit' of the curve to the data is shown by the sum of the squares of the difference as shown in Equation 6-5.

$$\chi^2 = |\mathbf{A} \cdot \mathbf{a} - \mathbf{b}|^2 \quad\text{..} (6\text{-}5)$$

The technique of singular value decomposition (e.g. Press *et al.*, 1986) can be applied to solve this equation. The SVD algorithm takes a matrix, in this case **A**, an $M{\times}N$ matrix (where $M{>}N$), and returns three matrices **U** ($M{\times}N$ column-orthogonal), **W** (diagonal $N{\times}N$), and **V** ($N{\times}N$ orthogonal), related to **A** by Equation 6-6.

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T \quad\text{..} (6\text{-}6)$$

The inverse of **A** is then trivial to calculate, except where the diagonal values in **W**, w_{ii}, are equal to zero. For zero (or small) values of w_{ii}, the columns of **V** form part of the nullspace of the equation (where $\mathbf{A} \cdot \mathbf{x} = 0$ for all **x**), and so their value in the inverse ($1 / w_{ii}$) can be set to zero. With this proviso, **a** is calculated as follows.

$$\mathbf{A} \cdot \mathbf{a} = \mathbf{b}$$
$$\mathbf{a} = \mathbf{A}^{-1} \cdot \mathbf{b}$$

$$\mathbf{a} = \mathbf{V} \cdot [diag(1/ w_{ii})] \cdot \mathbf{U}^T \cdot \mathbf{b} \quad\text{..} (6\text{-}7)$$

The SVD matrices will always give an answer to this equation, even when a solution did not exist for Equation 6-4. It can be proved (see Press *et al.*, 1986) that the values for **a** calculated from this final equation provide the solution with the least square error stipulation of Equation 6-5.

Discussion

There were occasions on which the fitted quadratic curve deviated from the original data. These situations most frequently occurred when the segment contained a sigmoid, or 's' shaped, contour. It would be possible to increase the number of parameters for the curve's equation so that a cubic curve could be used in place of a quadratic. The drawbacks to this approach are the increase in computational time (although this would still be much less than half a second), and the increase in complexity for comparing curves with cubic rather than quadratic equations.

An alternative solution would be to divide the curve into two sub-segments, each of which would then be fitted more closely by quadratic equations. This solution can then be repeated until the desired representational accuracy is achieved. Although this method has not been incorporated into the curve-fitting algorithm, since in general the curve-fitting segment to segment was deemed adequate, it should be a relatively simple modification to include in the software. Curves that require further subdivision may be identified by a high squared error value, specified in Equation 6-5 above. The division should ideally be made at a point that most closely divides the square of the error — in the case of an equal number of crossing points, at a point with a local maximum error; and for an odd number of crossing points, at one of the crossing points itself (i.e. zero error).

Another undesirable feature of this method was the jump between curves that would occur at the segment boundaries, where two sequential curves would not have the same frequency at their end/start point. Clearly, where this simple least squares error curve is used, a boundary error will occur unless the data corresponds to values taken from a series of quadratic functions. However, the standard deviations used in the singular value decomposition routine can be employed in such a way that those points at the beginning and end of a segment have a relatively smaller value of σ_i than those located near the centre. This would encourage the resultant curve to follow more closely the data points at the boundaries between segments, thereby resulting in a smoother transition. Since the majority of deviations from the curves tend to happen near a segment boundary (where the whistle is often changing its slope most rapidly), then using this weighting might result in a curve which has an overall higher error from the original data than would normally be desired.

Information on the whistle's intensity is discarded with this method, although some researchers have noted the possibility that the intensity might play a part in one dolphin identifying the whistles of another.

"Amplitude modulations within individual whistles were discussed by Lilly and Miller (1961), who noted that in most cases the amplitude reaches a peak, then falls during the rest of the emission. This tends to be true for whistles of only a single loop. Our measurements suggest that the lower frequencies usually contain greater energy than the upper frequency portions. On continuing repetitive multilooped whistles, this same pattern is repeated with the greater energy repetitively displayed in the lower frequencies. More rapid amplitude modulation may appear, especially in sections of rapid frequency modulation. As Tyack (1989) noted, very rapid amplitude modulations may appear as sidebands on spectrograms of whistles, and they may be perceived as introducing an unusual timbre to the whistle. Specific patterns of amplitude modulation, or the lack of them, are characteristic features of the signature whistle of each individual." (Caldwell *et al.*, 1990).

No studies have been conducted, however, on the role of internal whistle intensity changes on a whistle's information content, and this is not an attribute that has been included in recent studies. Brenda McCowan states

"Whistles are usually characterized by their relative change in frequency over time, known as whistle 'contour.' " (McCowan, 1995).

Furthermore, she attaches little significance to the role of intra-whistle intensity changes, and does not make any attempt to include them in her identification methods. Both of these positions seem consistent with colour spectrograms (representing intensity) of whistles recorded from dolphins in oceanariums and the Moray Firth. Contours *have* been observed in which rapid changes occur in intensity, which gives them a 'trilling' or 'bubbling' intonation. Whilst such clearly audible effects are probably deliberate, and indeed might hold some information, most whistles seem to be consistently more intense in the lower frequencies and quieter above approximately 15 kHz.

If studies of the role of intensity changes were deemed necessary, it would be quite simple to alter the existing algorithms to generate a second quadratic curve that would approximate the change in intensity with time across each whistle segment. Differences in the time-intensity contour could then be incorporated into quantifying

the similarity between two contours, and could be used in the calculation of the probability that they belong to the same whistle class.

Summary

Attempts by previous researchers to reduce the amount of data required to compare two whistle contours have been focused mainly in two areas. The first identifies features or attributes contained in the entire whistle, such as frequency range, start frequency, end frequency, time duration, and number of turning points (or 'loops'). The second one samples the contour at a reduced number of points along its length. Feature calculation and isolation has been proved to be effective for a small number of animals where *a priori* information on identity can be obtained, but when extended to high numbers of whistles from a larger but unknown number of individuals, it becomes a less useful analysis tool. Reduction of the number of sampling points reduces the information that may be analysed, but with no guarantee that it is preserving the features that are important for identifying the vocalising individual.

The technique chosen in this thesis for contour representation preserves the shape of the whistle by encoding it as a sequence of curved segments, each of which is simply labelled as rising, falling, or flat in frequency slope trends, or blank for a temporary break in the contour. This segment encoding of the target whistle allows rapid comparison with a previously held database, thus reducing the number of candidate whistles on which a more detailed comparison is needed. This more detailed analysis is based on the curves contained within the segments. The data storage requirements for this encoded form are very much less than for storing the long sequence of time-frequency-intensity values from the original spectrogram, and the preservation of the contour's shape ensures that the identity information contained in the signature whistle is not degraded.

This technique relies on the assumption that it is the shape of the time-frequency contour that contains the identity information, and all internal intensity information is discarded. It would, however, be possible to incorporate a second curve to fit the time-intensity contour of the whistle for each segment and for this to be used during further analysis.

Now that the data quantity in each contour has been reduced to a more manageable level, the next task is for quantitative comparisons between whistles based on the sequence of segment types and differences in the curves from one contour to another.

References

Caldwell, M.C., Caldwell, D.K., and Hall, N.R. (1970). "An experimental demonstration of the ability of an Atlantic bottlenosed dolphin to discriminate between whistles of other individuals of the same species." *Los Angeles County Mus. Nat. Hist. Found.* 6:35.

Caldwell, M.C., Caldwell, D.K., and Tyack, P.L. (1990). "Review of the signature-whistle hypothesis for the Atlantic bottlenosed dolphin." In *The Bottlenose Dolphin* (S. Leatherwood and R.R. Reeves, eds.), Academic Press, pp. 199–234.

Caldwell, M.C., Hall, N.R., and Caldwell, D.K. (1971). "Ability of an Atlantic bottlenose dolphin to discriminate between, and potentially identify to individual, the whistles of another species, the spotted dolphin." *Cetology* 6:1.

Lilly, J.C & Miller, A.M. (1961). "Sounds emitted by the bottlenose dolphin." *Science* 133: 1689–1693.

Mahadevan, S. (1994). "Identification of individual bottlenose dolphins by analysing their signature whistles." Master of Philosophy thesis, Loughborough University, United Kingdom. 116 pp.

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.

McCowan, B. & Reiss, D. (1995). "Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins (Delphinidae, *Tursiops truncatus*): a re-evaluation of the signature whistle hypothesis." *Ethology* 100: 194–209.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1986). *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, Chapter 12 Fourier Transform Spectral Methods, pp. 390–395.

Sturtivant, C.R. & Datta, S. (1996a). "Computer characterisation of dolphin whistles." European Cetacean Society 10th Annual Conference, Lisbon, Portugal, March 11th–13th, 1996.

Sturtivant, C.R. & Datta, S. (1996b). "Automated characterisation of dolphin whistles by computer." European Association for Aquatic Mammals 24th Annual Symposium, Albufeira, Portugal, March 15th–18th,1996.

Sturtivant, C.R, Datta, S., & Mahadevan, S. (1995). "Automated analysis of signature whistles of wild bottlenose dolphins (*Tursiops truncatus*) to determine group identity." Paper presented at the European Association for Aquatic Mammals 25th Annual Symposium, Nuremberg, 1995.

Tyack, P.L (1989). "Use of a telemetry device to identify which dolphin produces a sound: When bottlenose dolphins are interacting they mimic each others signature whistles." In *Dolphin Societies: methods of study* (K. Pryor and K.S. Norris, eds.), University of California Press, Berkeley.

Chapter 7: Whistle Classification

Introduction

There are two ways in which a pair of whistles may be compared using the current encoding. The first is to compare the overall shape of the two whistles using the sequence of segment types, and the second is to use the detailed shape by comparison of the quadratic curves which make up each of the matching segments. Some probability must be assigned to the similarity or dissimilarity of the two whistles in order to gain a quantitative result,. This probability may be calculated if some 'distance measure' between the two contours can be found, and a suitable probability distribution matched to the measure.

The simplest of measures would be to sum the differences in frequencies at corresponding times over the entirety of the two contours. This method can produce unwanted results, however. Using this method it is conceivable that two contours that had the same shape, but differed by a few hundred hertz, might be assigned a large distance measure, whereas two contours with similar average frequencies but whose shape was very different might be assigned a smaller distance measure.

This example encouraged the development of a hybrid distance measurement, which consisted of a combination of the differences between the curves in average frequency, frequency slope, and rate of change of frequency slope. In this way the relative importance of these three measurements of similarity can be tailored to each whistle class. Furthermore, the three component distances can be calculated directly from the quadratic parameters of the segments to be compared.

Clearly it is undesirable to calculate a distance measure between two segments with incompatible shapes, since the shape of the whistle is an important factor in our classification scheme. If we are to compare only those segments that are of the same type, a further distance measure must also be developed for differences in shape between two contours. For example, a 'flat' segment at a peak in a contour might be more significant in one class than in another, and its presence or absence in a target

contour should subsequently have markedly different effects on its eligibility for membership.

With no prior knowledge of the variability within a class of a whistle contour's shape, the comparison criteria need to be built up empirically as the number of contours presented increases. This task can fall victim to various classification problems due to the initial lack of data on which to build suitable generalised models. This uncertainty may lead to the assignment of two whistles to different classes when they should in fact be classified together, or vice versa. Thus there is a case for reclassification of whistle contours to solidify class membership criteria after a sufficient number of them have been analysed and assigned.

A technique utilising hidden Markov models (HMMs) was developed to provide a class membership probability based on the contour segment sequences. HMMs are a powerful pattern recognition tool that can be trained to selectively identify sequences of events. A description of how these models have been applied to the task of contour shape identification will be given.

The final probability that a contour belongs to a particular class can be calculated as the product of the probabilities from the segment similarity and the shape similarity. The segment similarity value can be found from the HMM for that class, which will give the probability that the class could produce the given whistle. The shape similarity can be found by calculating the average difference between the curves contained in the candidate contour and those contained in the class, for those segments that correspond between contours being compared.

Hidden Markov Modelling

Hidden Markov models have been a source of mathematical study since the late 1960s, but have not been used widely in pattern recognition analysis until the last decade. An excellent tutorial on HMM applications has been written by Lawrence Rabiner (Rabiner, 1989), which formed the basis for the HMM implementation used in this research. Some modifications to the basic implementation as laid out in Lawrence's paper were required, however, and these will be discussed. A brief introduction to the hidden Markov model will be presented in this section, with the

details of its application to whistle shape classification left until the next section. For a more detailed discussion of the theory behind the use of HMMs, the reader is directed to Lawrence's paper.

Markov Processes and the 'Hidden' Layer

A Markov process is a variant of a finite state automaton, consisting of a fixed number of states, transitions between states, and a set of output symbols (see Figure 7-1). At any one time the process can be in only one of the states, and it will then make a transition to another state (or back to itself). The state to which the process moves will depend upon the probabilities associated with the different transitions. If for any state, transitions are possible to all other states, then the process is termed 'ergodic'.
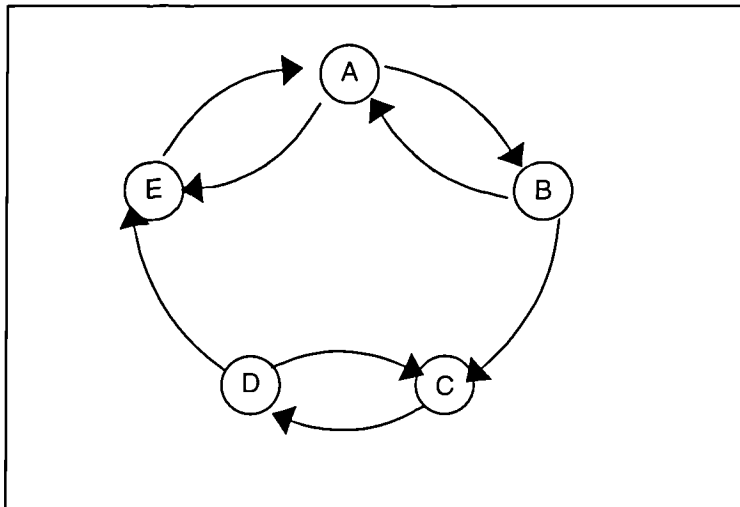
Figure 7-1: Representation of a (non-ergodic) Markov process

This system is quite easy to understand if for each state there is a unique output symbol, enabling us to follow the sequence of states by following the sequence of symbols. However, if we introduce a second probabilistic event such that the symbols output in each state are not unique, but correspond to a set of probabilities associated with that state, then the state sequence becomes unclear. Thus, the state of the system becomes the 'hidden layer' in the process, since this cannot be directly inferred from the observation sequence.

It is this combination of a finite number of states, transition probabilities from one state to another, and output symbol probabilities for each state, that forms the

hidden Markov process or model. The different components are often stated as follows:

- the set of states $S = \{S_1, S_2, \cdots, S_N\}$, for N states, and the state at time t as q_t;

- the set of output symbols $V = \{v_1, v_2, \cdots, v_M\}$, for M symbols;

- the state transition probability distribution $A = \{a_{ij}\}$, where

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \le i, j \le N \dots\dots\dots\dots\dots\dots(7\text{-}1)$$

(i.e. the probability that the next state is S_j given that the current state is S_i);

- the observation symbol probability distribution $B = \{b_j(k)\}$, where

$$b_j(k) = P[v_k \text{ at } t | q_t = S_j], \quad \begin{matrix} 1 \le j \le N \\ 1 \le k \le M \end{matrix} \dots\dots\dots\dots\dots\dots(7\text{-}2)$$

(i.e. the probability of outputting symbol v_k at time t given that the current state is S_j);

- The initial state distribution $\pi = \{\pi_i\}$, where

$$\pi_i = P[q_1 = S_i], \quad 1 \le i \le N \dots\dots\dots\dots\dots\dots(7\text{-}3)$$

(i.e. the probability that the process starts in state S_i).

The HMM can then be used to generate a sequence of observations O of length T, termed an observation vector, where $O = O_1 O_2 \cdots O_T$ and $O_t \in V$.

Determining Probabilities of Observation Sequences

In general, we will not want to use the HMM to generate a set of random observation sequences according to a set of probabilities. One alternative might be to manipulate the state transition and the observation symbol probability distributions so that the outputs of the HMM correspond to a specific class of whistles (i.e. the sequence of segment types). Then we might be able to calculate the probability that an observation sequence could have been generated by that HMM (the probability of the segment sequence given it belongs to that class).

The method of calculating this probability is not immediately apparent, since by its very nature we have no unique sequence of states through the HMM to generate any particular sequence of observations. The brute force method of solving the problem would be to calculate all possible state sequences through the HMM for T observations, and sum the associated probabilities for the symbols in the observation sequence. Since for an ergodic network, the number of state combinations that need to be considered would be N^T, this approach would appear singularly unappealing.

A solution requiring much less effort is known as the 'forward-backward' procedure (first described by Baum *et al.*, 1970). Two variables are calculated: α, known as the forward variable; and β, the backward variable. Their definitions are as follows:

$$\alpha_t(i) = P[O_1 \, O_2 \cdots O_t, \ q_t = S_i \mid \lambda] \quad \dotfill (7\text{-}4)$$

$$\beta_t(i) = P[O_{t+1} \, O_{t+2} \cdots O_T \mid q_t = S_i, \lambda] \quad \dotfill (7\text{-}5)$$

Here, λ is use to represent a specific HMM, so the forward variable $\alpha_t(i)$ represents the probability of observing a partial observation sequence up until time t, and at that point being in state S_i. Similarly, the backward variable $\beta_t(i)$ represents the probability of observing the remaining observations having reached time t in state S_i.

Only the forward variable needs to be used to calculate the probability of an observation sequence, and it can be constructed using the following algorithm. At time $t = 1$, $\alpha_1(i) = P[O_1, \ q_1 = S_i]$, representing the probability of starting in state S_i and observing the first output symbol. The probability of starting in any particular state is given by the initial state distribution π, therefore:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \le i \le N \quad \dotfill (7\text{-}6)$$

Values of α for $t > 1$ can be calculated inductively from previous values. To reach any position through the observation sequence, one must have reached the preceding position in some state, then have made the transition from that state to the current one, and also output the required symbol. This is shown in a mathematical form in Equation 7-7.

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) \, a_{ij} \right] b_j(O_{t+1}), \quad 1 \le t \le T-1 \quad \text{...............(7-7)}$$

Finally, when $t = T$, and all the observations have been included in that calculation, α_T will contain the probability that after all the observations have been made the process has ended in each of the N states. The total probability of observing the sequence can be calculated by summing all the α_T values over all available states (Equation 7-8).

$$P[O|\lambda] = \sum_{i=1}^{N} \alpha_T(i) \quad \text{...(7-8)}$$

Thus, the probability that a particular observation sequence is produced by a HMM can be calculated in $T \times N$ steps, rather than the N^T steps of a brute force approach. If we use an observation sequence that corresponds to a segment sequence, then this result corresponds to the probability that a target sequence of segments would match a randomly produced segment sequence of the same length from that class.

Training a Hidden Markov Model

The evaluation of the class membership probability for an observation sequence presumes that in some way the HMM has been manipulated to produce sequences which correctly represent the members of that class. The technique that allows HMM modifications that increase this probability makes use of the previously unused 'backward' variable β. The procedure laid out by Rabiner (Rabiner, 1989) will be described here, although modifications to this algorithm were necessary.

The backwards variable may be calculated in an iterative way similar to that used for α. Clearly, at time T the observation sequence has been completed, so all values for β will be unity (from Equation 7-5). Preceding values can be calculated from their successors by considering all of the possible future states, as shown in Equation 7-9.

$$\beta_T(i) = 1, \qquad\qquad 1 \le i \le N$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} \, b_j(O_{t+1}) \, \beta_{t+1}(j), \qquad \begin{matrix} 1 \le t \le T-1 \\ 1 \le i \le N \end{matrix} \quad \text{........................(7-9)}$$

Training the HMM occurs by modifying the variables so that they reflect better the probabilities of producing each observation in the sequence. For this to happen, both the state transition and output symbol probability distributions, and the initial state distribution are altered. The procedure is aided by the use of a variable ξ, which describes the probability of a transition from one state to another at a particular time based on an observation sequence. Its derivation is shown in Equations 7-10, 7-11, and 7-12.

$$\xi_t(i,j) = P[q_t = S_i, q_{t+1} = S_j \mid O, \lambda] \quad\text{........................}(7\text{-}10)$$

$$= \frac{\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)}{P[O\mid\lambda]} \quad\text{..........................}(7\text{-}11)$$

$$= \frac{\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)} \quad\text{......................}(7\text{-}12)$$

The numerator in Equation 7-12 evaluates the probability of reaching state S_i, then making the transition from there to state S_j and outputting the correct symbol, and then reaching the end of the observation sequence. The denominator calculates the summed probability for all state transitions at time t, thus converting ξ into a probability measure.

A second variable γ is defined as the probability of being in a state at a specific time, given an observation sequence (Equation 7-13).

$$\gamma_t(i) = P[q_t = S_i \mid O, \lambda] \quad\text{..................................}(7\text{-}13)$$

This probability can be derived from ξ by summing over the second state S_j, i.e. the probability that we are in state S_i and a transition to some other state is made (Equation 7-14).

$$\gamma_t(i) = \sum_{j=1}^{N}\xi_t(i,j) \quad\text{...}(7\text{-}14)$$

The two variables may now be used to re-estimate the three probability distributions for the HMM as shown in Equations 7-15, 7-16, and 7-17.

$$\tilde{\pi}_i = \text{expected probability in state } S_i \text{ at time } t = 1$$
$$= \gamma_1(i) \qquad\qquad\qquad \dots\dots\dots\dots(7\text{-}15)$$

$$\tilde{a}_{ij} = \frac{\text{expected number of transitions from state } S_i \text{ to } S_j}{\text{expected number of transitions from state } S_i}$$

$$= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \qquad\qquad \dots\dots\dots(7\text{-}16)$$

$$\tilde{b}_i(k) = \frac{\text{expected number of times in state } S_i \text{ and observing symbol } v_k}{\text{expected number of times in state } S_i}$$

$$= \frac{\sum_{\substack{t=1 \\ \text{s.t. } O_t = v_k}}^{T-1} \gamma_t(i)}{\sum_{t=1}^{T-1} \gamma_t(i)} \qquad\qquad .(7\text{-}17)$$

Thus the three re-estimated probabilities $\tilde{\pi}, \tilde{a},$ and $\tilde{b}$ will produce a more accurate response to any particular observation sequence than $\pi, a,$ and b.

Training for More than One Observation Sequence

If each HMM is to represent a class of whistles, training it to recognise only one example of that class is clearly inadequate. In this case we have a set of observation sequences, and the desired result is to produce a HMM with the highest product for the probabilities of these sequences. Rabiner (1989) suggests the following routine for combining multiple observation sequences when training a left-right or Bakis HMM (where no transitions to a lower state is allowed). This can be generalised so that it can be applied to all HMMs with the inclusion of the π re-estimation equation.

For a set of K observation sequences $\mathbf{O} = \{\mathbf{O}^{(1)}, \mathbf{O}^{(2)}, \cdots, \mathbf{O}^{(K)}\}$,

$$P[\mathbf{O}\,|\,\lambda] = \prod_{k=1}^{K} P[\mathbf{O}^{(k)}\,|\,\lambda]$$
$$= \prod_{k=1}^{K} P_k \qquad\qquad \dots\dots\dots\dots\dots\dots(7\text{-}18)$$

And the re-estimation formulae are modified as follows:

$$\tilde{\pi}_i = \sum_{k=1}^{K} \frac{1}{P_k} \alpha_1^k(i)\,\beta_1^k(i)$$

$$= \sum_{k=1}^{K} \frac{1}{P_k} \gamma_1^k(i) \qquad \dots\dots\dots\dots\dots\dots\dots (7\text{-}19)$$

$$\tilde{a}_{ij} = \frac{\displaystyle\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i)\, a_{ij}^k\, b_j^k(O_{t+1}^{(k)})\, \beta_{t+1}^k(i)}{\displaystyle\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i)\, \beta_t^k(i)}$$

$$\qquad\qquad\qquad\qquad \dots\dots\dots\dots\dots\dots (7\text{-}20)$$

$$= \frac{\displaystyle\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \xi_t^k(i,j)}{\displaystyle\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \gamma_t^k(i)}$$

$$\tilde{b}_j(l) = \frac{\displaystyle\sum_{k=1}^{K} \frac{1}{P_k} \sum_{\substack{t=1 \\ \text{s.t. } O_t^{(k)}=v_l}}^{T_k-1} \alpha_t^k(j)\, \beta_t^k(j)}{\displaystyle\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(j)\, \beta_t^k(j)}$$

$$\qquad\qquad\qquad\qquad \dots\dots\dots\dots\dots\dots (7\text{-}21)$$

$$= \frac{\displaystyle\sum_{k=1}^{K} \frac{1}{P_k} \sum_{\substack{t=1 \\ \text{s.t. } O_t^{(k)}=v_l}}^{T_k-1} \gamma_t^k(j)}{\displaystyle\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \gamma_t^k(j)}$$

These formulae can be applied iteratively to the HMM until the change in the model has achieved a maximum probability for Equation 7-18. This method guarantees to find a probability maximum given the observation sequence set, but this might only be a local maximum and is not necessarily the desired global probability maximum. No technique has thus far been discovered that will always assure the global maximisation of this equation.

Segment Recognition

The production of a sequence of observations to represent a whistle contour was a simple task, since the labels used in the contour's segment list could be mapped numerically as the observation symbol set for the HMM. Since the observation sequences were, however, of differing lengths even within the same whistle class, an

93

additional symbol was required in order to indicate that the contour ended at that point. An example will serve to explain why this extra symbol was required. If a HMM was trained to represent an 'up, flat, down' segment sequence, it would also similarly represent all contours which contained some initial part of that sequence. It can be seen from Equations 7-7 and 7-8 that if the initial part of a training sequence is presented to the HMM, its probability will be greater than or equal to that of the complete training sequence, since a_{ij} and $b_j(k)$ are always less than or equal to one. Thus the sequences 'up, flat' and simply 'up' would then satisfy the criteria for class membership, since no 'termination' state exists in this model. Rather than producing a fixed termination state, which would have complicated the algorithms, a termination symbol was introduced into the observation symbol set. With this inclusion, a HMM trained to represent 'up, flat, down, end' would not also incorrectly represent 'up, flat, end'.

The number of hidden states in the HMM needed some consideration. The number of segments into which a whistle was split varied greatly between whistle types, so no fixed value naturally suggested itself. The value that was chosen for the number of hidden states was the number of segments in the whistle that was first assigned to the class, plus an extra state for the termination symbol. In this way the output symbol probability distribution for each state could be trained to represent the segment type at a specific position in the observation sequence of the single contour in the class, and subsequent contours will modify the probabilities in this initial template. One of the advantages to this method is that in using the smallest possible number of states, the problem of over-training the HMM can be overcome, where it would recognises several individual sequences instead of generalising features.

Modifications to the HMM Algorithm

Whilst implementing the re-estimation algorithms detailed in Rabiner (1989) for a single observation sequence, it was discovered that the re-estimation formula for the output symbol probability distribution as described would not include all the symbols contained in the training sequence. Specifically, the terminal symbol of the sequence did not affect any of the values in the next iteration. Examination of the algorithm indicated why this was the case. Equation 7-17 suggested that for γ only values from

1 to $T-1$ should be used when re-estimating the output symbol probability variable b, and so the check for $O_T = v_k$ specifically (for the last symbol in the sequence) is never made. A minor alteration of the equation, which does include the final symbol, is shown in Equation 7-22.

$$\tilde{b}_i(k) = \frac{\sum_{t=1 \atop \text{s.t. } O_t = v_k}^{T} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)} \quad\quad\quad\quad\quad (7\text{-}22)$$

However, the definition for γ derives from values calculated for ξ, which are only defined for the observation range from 1 to $T-1$. Thus, γ is also only valid for observations between 1 and $T-1$ (Equation 7-23).

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) = \sum_{j=1}^{N} \frac{\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)}, \quad 1 \le t \le T-1 \ \dots (7\text{-}23)$$

This problem was solved by calculating values for $\gamma_T(i)$ separately from those for previous observations, which still used the values for ξ in their calculation. From the iterative equation for β (Equation 7-9):

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)$$

Thus, $\quad \gamma_t(i) = \dfrac{\alpha_t(i) \sum\limits_{j=1}^{N} a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)}{\sum\limits_{i=1}^{N}\alpha_t(i)\sum\limits_{j=1}^{N} a_{ij}\, b_j(O_{t+1})\,\beta_{t+1}(j)}$

$$= \frac{\alpha_t(i)\,\beta_t(i)}{\sum\limits_{i=1}^{N}\alpha_t(i)\,\beta_t(i)} \quad\quad\quad\quad\quad (7\text{-}24)$$

For $t = T$, $\quad \gamma_T(i) = \dfrac{\alpha_T(i)}{\sum\limits_{i=1}^{N}\alpha_T(i)}, \quad$ since $\beta_T(i) = 1$. $\dots$ (7-25)

With this alteration, the equations performed correctly, utilising all of the observations in a sequence. It should be noted that other papers also contain this error

95

(e.g. Picone, 1990), although the use of α and β equivalents in place of γ and ξ throughout the calculations overcomes the problem at the expense of some additional complexity.

Multiple Segment Sequences

The equations for re-estimating the probability distribution variables for multiple observation sequences (Equations 7-19, 7-20, and 7-21) weight more heavily those sequences that are currently less probable by using the factor $1/P_k$ in the calculations. When these equations were implemented, and small numbers of sequences were used as a training set, then observation of the changes in probability between consecutive iterations indicated that two observation sequences would quickly become more probable that the rest. The probabilities for these two sequences would then begin to oscillate between progressively higher and lower values. It was deduced that the use of weightings of the inverse of the probability produced instability in the equations.

A different weighting was investigated for the sequences, maintaining the idea that those observation sequences that were less likely should be weighted more heavily during subsequent re-estimation than those that were already well represented by the current HMM structure. A simple negation reversal of the current probability (i.e. subtraction from one) sufficed to produce stable behaviour in the re-estimation iterations, whilst still allowing rapid convergence in well under a second. The updated re-estimation equations that were used in the HMM implementation are shown in Equations 7-26, 7-27, and 7-28.

$$\tilde{\pi}_i = \sum_{k=1}^{K} (1 - P_k)\, \gamma_1^k(i) \qquad \text{...(7-27)}$$

$$\tilde{a}_{ij} = \frac{\displaystyle\sum_{k=1}^{K} (1 - P_k) \sum_{t=1}^{T_k-1} \xi_t^k(i,j)}{\displaystyle\sum_{k=1}^{K} (1 - P_k) \sum_{t=1}^{T_k-1} \gamma_t^k(i)} \qquad \text{...(7-28)}$$

$$\tilde{b}_j(l) = \frac{\sum_{k=1}^{K}(1-P_k)\; \sum_{\substack{t=1 \\ \text{s.t. } O_t^{(k)}\; v_l}}^{T_k}\gamma_t^k(j)}{\sum_{k=1}^{K}(1-P_k)\sum_{t=1}^{T_k}\gamma_t^k(j)} \quad\text{...(7-29)}$$

The Distance between Two Curves

The contour's shape within each of the segments was approximated by a quadratic curve, and hence the comparison of any two segments from different whistles becomes the comparison of the equations. The distance measure based on curve differences was formed from three parts: differences in frequency; differences in contour slope; and differences in rate of change of slope. The three attributes that were measured were presumed to be normally distributed (Gaussian) for each segment, an assumption developed by visual inspection of these attributes for a number of whistles. The difference measures were therefore formed from the difference of two random variables from normal distributions that had the same mean and standard deviation, and hence the measures would also be normally distributed around a mean of zero.

The standard deviations for each of the three difference measures were calculated by comparison of all whistles in a class against each other (see Equation 7-36 later in this chapter). The similarity to the class of any candidate whistle contour could subsequently be quantified by comparing the candidate contour with those already in the class, and calculating the three distance measures (δ_0, δ_1, and δ_2 in Equations 7-33 to 7-35) according to Equation 7-30. These values could then be employed to calculate three associated degrees of similarity (Equation 7-31) by the use of the normal distribution, and the average of the three values calculated to indicate the overall similarity to the class based on segment curves (Equation 7-37).

$$\tilde{\delta}_i = \frac{\sum_{\text{all whistle pairs}}\left(\sum_{\text{all matching segments}}\delta_i\right)}{\text{Total number of matching segments}} \quad\text{.............................(7-30)}$$

$$s_i = P[\delta_i > \delta \mid \delta \text{ is } N(0,\sigma_i)] \quad\text{...(7-31)}$$

These similarity values represent the probability that any randomly chosen whistle from the class will have its distance measure greater than that of the candidate whistle. The overall value therefore gives the average probability of a whistle from the class having one of the three distance measures (randomly chosen) greater than those of the candidate whistle.
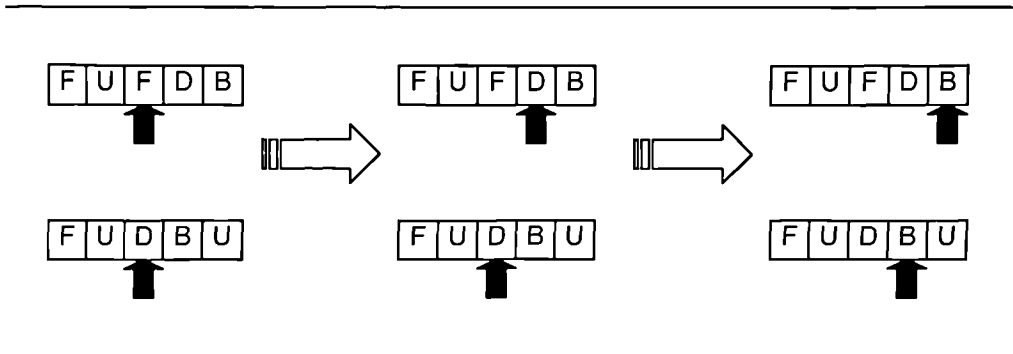

Figure 7-2: Example of algorithm for moving over 'extra' segments

Segment Matching

Only segments of the same type were used in the comparison calculations. This prevented obviously dissimilar segments from producing a disproportionate effect on the mean distance measure, since segment discrepancies were already accounted for in the HMM calculation of segment similarity. A two segment look-ahead system was used in order to keep the comparisons in step when an extra segment was present within one contour but not the other. Whenever segments did not match, the next segment in one contour was considered for comparison with the original in the second, and if a match had still not been found the next segment in the second contour was considered with the original in the first. The comparison would go to a second stage if no match were found, whereupon the next-but-one segments would be considered. (See Figure 7-2.) If a match were still missing at this second look-ahead, both segments were discarded, and comparison would move on to the next pair of segments.

Occasionally, there would be no match at all between two entire contours, and no distance measure was possible. When this occurred, the similarity measure was set to zero for convenience, since any other value would be misleading.
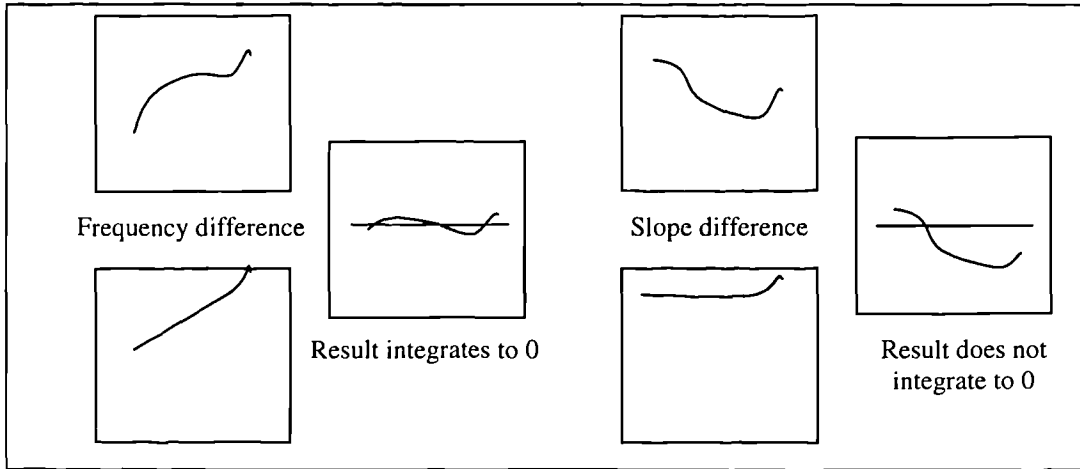
Figure 7-3: Explanation of why differences that average out in one measure will not in the next

The Three Difference Calculations

At its simplest level, the difference between curves contained within two segments might be calculated as the integral of the difference between them over the length of the segment. Since we are using three distance measures on the segments, we do not need to square the differences in the calculations. We may do this since differences that become averaged out to zero in one measure (e.g. frequency difference), will not be averaged out in the next (e.g. slope difference) as shown in Figure 7-3. Similarly, the same thing with averaging errors would occur in the slope difference measure, which would be carried through to the rate of change of slope distance measure. This is due to the three quadratic parameters being also represented in the three difference equations in an orthogonal manner. The form of the difference equation is shown in Equation 7-32.

$$\delta = \int_{t=0}^{T} \left(f_1(t) - f_2(t) \right) dt \quad \text{..............................(7-32)}$$

One problem that is immediately apparent in this equation is that of which value for T should be used, since two segments from different whistles will rarely be of the same length. Two options suggested themselves — either to normalise the length of the segments (similar to the technique of 'dynamic time warping' [Sakoe and Chiba, 1978]), or to choose the longer or shorter of the two segment durations. If the segment was to have its length altered to normalise the durations, some account should be taken of this modification in the resultant probability measure. Since no information was available on the distribution of segment durations per individual

99

dolphin, the time for the integral was taken over the longer of the two segments. This method ensures that all of the two segments were compared, no matter which of the two segments was the longer, whereas if the shorter time were chosen for the integral, some of one segment would not be included in the calculation. The curve values for the additional time period for the shorter segment are easily available from the quadratic equation, and can be thought of as 'what would have happened if the shorter segments had carried on' for the purposes of the comparison.

The difference in rate of change of slope was the easiest of the three calculations to make, since this could be derived directly from the highest order quadratic parameter, a_2. For the two quadratic curves, $f(t)$ and $g(t)$, the measure was calculated as follows:

$$\text{For } f(t) = a_0 + a_1 t + a_2 t^2,$$

$$g(t) = b_0 + b_1 t + b_2 t^2$$

$$\begin{aligned} \delta_2 &= \int_{t=0}^{T} (a_2 - b_2)\, dt \\ &= T(a_2 - b_2) \end{aligned} \quad\text{.......................................(7-33)}$$

Calculations for the lower order distances become more complicated, but can be derived in a similar fashion (Equations 7-34 and 7-35).

$$\begin{aligned} \delta_1 &= \int_{t=0}^{T} \left(f'(t) - g'(t) \right) dt \\ &= \int_{t=0}^{T} \left((a_2 T + a_1) - (b_2 T + b_1) \right) dt \\ &= T\left(\frac{T}{2}(a_2 - b_2) + (a_1 - b_1) \right) \\ &= T\left(\frac{\delta_2}{2} + (a_1 - b_1) \right) \end{aligned} \quad\text{...(7-34)}$$

$$\begin{aligned} \delta_0 &= \int_{t=0}^{T} \left(f(t) - g(t) \right) dt \\ &= \int_{t=0}^{T} \left((a_2 t^2 + a_1 t + a_0) - (b_2 t^2 + b_1 t + b_0) \right) dt \\ &= T\left(\frac{T^2}{3}(a_2 - b_2) + \frac{T}{2}(a_1 - b_1) + (a_0 - b_0) \right) \\ &= T\left(T\left(\frac{T(a_2 - b_2)}{3} + \frac{(a_1 - b_1)}{2} \right) + (a_0 - b_0) \right) \end{aligned} \quad\text{............................(7-35)}$$

These three distance measures were calculated for each pair of whistles contained in the class, and then an estimate for the standard deviation for each measure was calculated. Equation 7-36 details the equation for calculating the standard deviation estimate for a distance measure. No calculation for the mean was necessary in the equation, since its value should be zero.

$$\sigma_i = \sqrt{\frac{1}{\text{Number of segments} - 1} \sum_{\text{All segments}} (\delta_i)^2}, \quad 0 \le i \le 2 \quad \text{................(7-36)}$$

The degree to which a whistle is similar to a class can then by calculated in one of three ways using the three distance measures and the calculated standard deviations. The probabilities associated with the distance measure can be calculated using a Gaussian distribution. Each probability signifies the degree to which the whistle is representative of the entire class using that specific measure. Hence, an overall probability can be found by calculating the average of the three probabilities.

$$\text{degree of similarity} = Ave\left(P[\delta_0 \mid N(0,\sigma_0)], \, P[\delta_1 \mid N(0,\sigma_1)], \, P[\delta_2 \mid N(0,\sigma_2)]\right) \text{....(7-37)}$$

The degree of similarity of a candidate whistle with a class is calculated by calculating the sum of the similarity measures of the segments within the candidate whistle with those corresponding ones in all whistles already contained within the class, and then calculating the average.

Bayesian Theory

The results from the curve comparison indicated the degree to which the curves contained within a whistle's segments were representative of a given class. Similarly, the HMM results contained the probability that the specified sequence of segments contained in the whistle were formed, given that it belonged to a particular class. However, the value that we require is the probability that a whistle belongs to a class, given the whistle's curve or segment parameters. Fortunately, this conversion can be accomplished using Bayes's theory for conditional probabilities.

We can combine the similarity and probability measures together to form the joint probability simply by finding their product, since this does not change the conditional side of the probability (Equation 7-38).

$$P[\text{whistle} \mid \text{class}] = P[\text{whistle segments} \mid \text{class}] \times \left(\text{whistle similarity}\right) \dots\dots\dots(7\text{-}38)$$

From Bayesian theory, the probability of a class given a whistle can be calculated from the probability of a whistle given a class, provided that information on the probability of a class and the probability of a whistle's occurrence can be found (Equation 7-39).

$$P[\text{class} \mid \text{whistle}] = \frac{P[\text{whistle} \mid \text{class}] \times P[\text{class}]}{P[\text{whistle}]} \dots\dots\dots\dots\dots\dots(7\text{-}39)$$

The (unconditional) probability of a whistle can be found by summing the conditional probabilities for that whistle over all possible classes multiplied by the probability of that class (Equation 7-40). The probability of the class can be found as the number of whistles contained within a class as a proportion of the total number of whistles classified (Equation 7-41).

$$P[\text{whistle}] = \sum_{\text{all classes}} P[\text{whistle} \mid \text{class}] \, P[\text{class}] \dots\dots\dots\dots\dots\dots(7\text{-}40)$$

$$P[\text{class}] = \frac{\text{No. of whistles in class}}{\sum_{\text{all classes}} \text{No. of whistles in class}} \dots\dots\dots\dots\dots\dots(7\text{-}41)$$

Classification Decision

The probability value for the whistle to be a member of each of the classes was calculated by the above routines. The question still remained, however, as to what threshold is suitable for a whistle to be included in an already existing class, or for it to form the basis for a new class. At an early stage, when each class contains only a few contours, the probability measures will be quite uncertain. In this case intervention by an experienced user would aid initial classification, or alternatively one could err on the side of caution and classify separately all whistles which fall outside some membership minimum probability score. Later, once more information about the spread of a class's contours has become more certain, they could be merged together.

When a candidate whistle is tested against the set of current classes, a probability of class membership is calculated for each class. The Bayesian probability

correction requires the probability of occurrence of each class which, if we base this on the number of whistles already allocated to existing classes, will result in the sum of all class probabilities reaching 100%. Clearly the calculations must be modified so that the probability that a whistle might fall into a new class can be reflected. The whistle could then be assigned to whichever class presented the highest probability score — either assignment to an existing class, or for the whistle to form the basis for a new class. If the whistle is placed into an existing class, the class parameters (i.e. HMM behaviour and measurement standard deviations) may then be updated to include this new member.

Accounting for Uncertainty in the HMM

A probability score near 100% was often achieved when a HMM was trained with just one observation sequence. If left unmodified, this value indicates a vanishingly small probability that that class could encompass any whistle with a different segment sequence. Some indication of the uncertainty contained within the class is required, but since no *a priori* knowledge is available about the whistle distribution within each class, an assumption had to be made as to its value. For a single whistle class, the whistle was found to match approximately 30% of the whistles within the final class, from previous observations, leaving 70% of the whistles with uncertain sequences. An additional whistle would likewise account for 30% of the remaining 70% of sequences, etc. Thus the proportion of whistles u contained in the class that were still not represented by an existing whistle member could be found using Equation 7-42, where N is the number of whistles assigned to that class.

$$u = (0.70)^N \quad \text{..} \quad (7\text{-}42)$$

The way in which the original HMM should be modified to take into account this uncertainty figure was deduced by considering the final HMM to be the combination of two smaller ones — the original HMM, and a 'flat' HMM. This second HMM is made 'flat' by setting all symbol outputs and transitions to have the same probability, and is of the same geometry as the first. The combination is made by having a probability ε of making a transition between the two. No output symbol is produced when the current state moves between the two HMMs.

The algorithm was implemented by modifying parameters of the original HMM with those of the flat HMM using ε as the weighting (Equations 7-43, 7-44, and 7-45). Since from the definition for α (from Equations 7-6 and 7-7) the factor ε will be included in the calculation of α twice, the required value for ε is calculated from the uncertainty value u according to Equation 7-46.

$$\pi'_t = (1-\varepsilon)\pi_t + \varepsilon\,\pi_t^{(f)} \quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(7\text{-}43)$$

$$a'_{ij} = (1-\varepsilon)a_{ij} + \varepsilon\,a_{ij}^{(f)} \quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(7\text{-}44)$$

$$b'_{jk} = (1-\varepsilon)b_{jk} + \varepsilon\,b_{jk}^{(f)} \quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(7\text{-}45)$$

$$\varepsilon = \sqrt[2]{u} \quad\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(7\text{-}46)$$

When a further observation sequence was placed in the class, the HMM was modified to include this observation based on its previous values, rather than calculating a new HMM from untrained parameters. Since the training algorithm was guaranteed only to find a local maximum to the probability scores rather than the optimum global one, resetting the parameters to an untrained state might have caused the trained HMM to settle into another, less representational, state. Although it is possible that this new configuration might produce better probability values, this is by no means guaranteed, and it was decided that the instability of the results would adversely affect the classification decision process. The inclusion of the uncertainty modifications in the HMM parameters allowed the retraining process to have an extra degree of flexibility in finding a new local maximum without altering the parameters too far from the maximum previously calculated.

Accounting for New Classes

The Bayesian equations for converting the probability values to those required for class membership did not take into account a 'new' whistle class, thus the sum of all the existing class probabilities would be 100%. An additional 'new class' was incorporated into the Bayesian calculations, and a fixed membership probability (for a given class) of 25% was used with a class probability a straight proportion of the number of classes currently in existence (for use with Equation 7-41). This had the

effect of indicating that a candidate whistle should be placed in a new class when all membership probabilities for other classes were low.

The class probability value for the 'new' class was set to the reciprocal of the current number of classes identified plus one, and would thus be 100% with no classes identified, 50% with one, 33% with two, etc. This calculation incorporates the decreasing probability that a new class will be discovered as the number of classes already identified increases. Since no information is available about the number of whistle types, we are forced to make an approximation for the calculations. Although the 'new' class probability decreases quite rapidly, this can be justified since the different whistle types will not be presented randomly since the dolphins will be recorded in small groups. So if one whistle type is assigned to a class, it becomes more probable that that whistle type will again be presented within the next few whistles.

As the database of classified whistles increases, and the parameters for membership of each class become more well defined, the probability associated with the 'new' class become less important, since a new whistle type will have very small membership probabilities with any existing class. After Bayesian correction, the 'new' class probability would then be close to 100%.

Discussion

The problem of a suitable initial state for an 'untrained' HMM should be mentioned. It was presumed that the most easily trainable HMM would be one in which all state transitions and output symbol probabilities were equal. However, an HMM in this state could not be trained at all by the techniques outlined in this chapter, since there would be no sequence of states more likely to produce the observation sequence than another. A useful metaphor to explain this situation is that of a ball that is perfectly balanced on top of a hill, and is thus unable to roll downwards to find a more 'natural' position. A suggestion made by Rabiner (Rabiner 1989) is to perturb the initial state distribution so that the probabilities become unequal, thus enabling one route to predominate over the rest. Of course, this must also be accompanied by perturbation of the observed symbol probability distribution, or otherwise all routes will still remain equally probable.

Even when random small movements from the equilibrium position were added to the untrained HMM, the training procedure would often give very variable results. Training consisted of a number of re-estimations based upon the observation sequences, and terminated when no further increase in the joint probability of the observations was forthcoming. Successive training from an initial perturbed untrained HMM produced fully trained probabilities which varied widely from less than 1% to close to 100%, with several repeatedly observed local maximums between these two values.

The number of states available in the model had an effect on the fully trained probability value, since a non-repeating whistle containing 5 segments could not be fully represented by fewer than 6 states (the extra one due to the requirement for an 'end' observation). The use of a larger number of states than there were segments in the observation sequence simply increased the calculation time with no concomitant increase in representational accuracy. Thus, if an observation sequence containing T observations formed the basis for a new class, the HMM used to represent that class was assigned $T+1$ states.

The problem of encountering a poor local maximum in the training procedure could then be solved by pre-defining a more probable sequence of state transitions through the HMM, such that each state on this sequence was more likely to produce the required output symbol at that time. Convergence of the training procedure to 100% was subsequently very rapidly and reliably achieved.

A known problem with pattern recognition techniques which involve a learned response occurs due to the observation sequences themselves being recognised, rather than the more general properties contained within those sequences. For example, a recognition routine with a sufficiently large storage capacity might be trained with three utterances of the word 'horse'. The routine might learn to recognise the entirety of each of the three utterances rather than the attributes of the word itself, and would thus incorrectly reject a fourth utterance of the same word. Since the number of states for our HMMs was restricted to the number of observations within each sequence, two or more individual whistle contours were incapable of simultaneous explicit representation, and so recognition was limited to patterns contained within the group as a whole.

Summary

The similarity between whistles was measured using two components:

- The overall 'shape' of the whistle (the sequence of frequency rises, falls, flats, and gaps in the contour);

- The difference in the frequency, frequency slope, and rate of change of slope, for areas between pairs of whistles that corresponded in their 'shape' structure.

The 'shape' of the whistle was defined with regard to the sequence of segment types (as described in the previous chapter). The spread of the segment type sequences within a whistle group was represented using a hidden Markov model (HMM). By restricting the number of states in use by the model, the similarities between segment sequences were learned instead of the individual sequences themselves. The model provided a measure of the probability that any sequence of segments contained by a whistle could have been produced by the current parameters of that class. The number of whistles that had not yet been included in the class was taken into account in the calculation of the shape probability, using an estimate based upon visual inspection of the order of presentation of whistle series from tapes.

The second probability measure consisted of a more detailed comparison of the structure of the whistles within matching segments. A segment-matching algorithm was developed that was tolerant of up to two consecutive additional segments or missing segments within each whistle. When two segments matched, differences were calculated for frequency, frequency slope, and rate of change of slope, all based on the quadratic parameters previously fitted to the whistle contour for those segments. By visual inspection, these three parameters were observed to follow a Gaussian distribution, and therefore the difference between two occurrences of a parameter will also form a Gaussian distribution with a mean of zero. From this information, the standard deviation for each of the three difference measures can be calculated for any class containing a large enough number of whistles. Measurements of the average differences in the parameters between a candidate whistle, and whistles contained in the class, can then be used to calculate the similarity of the whistle's detailed shape with those contained within that class. In circumstances where the

class contains only a few whistle members, an estimate of the standard deviations for the differences was made based on previously observed classes.

The two probabilities for shape and structure were combined by multiplication, and Bayes's theorem on conditional probabilities was used to convert from the probability of the whistle being produced by a specified class, to the probability that a class contains the given whistle. This required estimates of class probability (including that for the formation of a new class) and the probability of the whistle being observed. Since no *a priori* information was available, estimates were made based on visual inspection of sequences of whistles sampled from a tape.

The final probabilities for class membership were used either to assign the whistle to an existing class, or to form a new class based on that whistle's parameters, depending on which result had the highest probability. Although not implemented in the software, after some time it might be desirable to reconsider the classes to which each whistle belongs, since the parameters defining each class will have become better defined.

References

Baum, L.E., Petrie, T., Soules, G., and Weiss, N. (1970). "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains." *Annals of Mathematical Statistics* **41**: 164–171.

Picone, J. (1990). "Continuous speech recognition using hidden Markov models." *IEEE ASSP Magazine, July 1990*: 26–41.

Rabiner, L. R. (1989). "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* **77**(2): 257–285.

Sakoe, H., and Chiba, S. (1978). "Dynamic programming algorithm optimization for spoken word recognition." *IEEE Trans. Acoust. Speech Signal Processing* **26**:43-49.

Chapter 8: Results

Introduction

The classification techniques allowed whistles to be incorporated into a pre-existing database of whistles. An indication could also be made by the software (based on quantified probabilities) that any whistles should belong in a new class that does not currently exist in the database. A class based on a small number of whistles does not, however, provide statistics about itself that can be treated with very much confidence. Since no *a priori* knowledge was available about class or whistle frequencies of occurrence, some assumptions based on empirical evidence were required.

In order to prove the software behaved as designed, data from typical recordings of wild dolphins was required, which could then be extracted, encoded, and classified, and the results analysed for differences between recording times. Data was taken from recordings made in the north Atlantic during trawl trials, and also from three days of recordings during 1992 in the Moray Firth.

Whistle Structure Standard Deviation

As noted in the preceding chapter, when a class contains a single whistle as its member, no information on the standard deviation of the three distance measures can be determined. Since this information was essential if the distance measures between a candidate whistle and the previously classified whistle were to be converted into a probability score, it was necessary to calculate their values empirically.

The data required for this task was gathered from whistles from three dolphins recorded at Kolmårdens Djurpark, Sweden. The contours for 26 whistles were isolated, encoded, and entered into the classification algorithms, with any apparent erroneous classifications corrected manually. The resulting means for the distance measures and their standard deviations were calculated for each class, and are presented in Table 8-1.

Dolphin name	No. whistles (segments)		Frequency measure	Slope measure	Slope change rate
Sisu	3 (14)	mean	1406	911	6415
		s.d.	312	1065	30140
		% s.d.	22.2%	117%	470%
Vindi	9 (42)	mean	7574	2236	-6516
		s.d.	242	1726	12048
		% s.d.	3.20%	77.2%	185%
Nephele	14 (176)	mean	1060	2050	8151
		s.d.	464	2433	23029
		% s.d.	43.8%	119%	283%

Table 8-1 — Means and standard deviations of distance measures for three whistle classes

The table shows the standard deviations for the three whistles both as absolute values and as percentages of the mean measures. The dolphin 'Sisu' had only three whistles recorded during the period consisting of fourteen segments, and so the standard deviations for the measures may not be as reliable as for the other two whistles. Typical contours from the three dolphins are shown in Figure 8-1, and it can be seen that each has a very different contour from the other two.

Vindi's whistle was the simplest of the three, with a long flat main contour sometimes preceded by a short rising period and short, shallow fall at the end. Table 8-1 indicates the strong frequency stability of the whistle by the low value for the standard deviation of the frequency distance measure. This whistle also exhibits the lowest standard deviation for the other two measures.

Nephele often produced a series of similar whistles with short breaks of approximately one quarter to one half a second between them (not shown in Figure 8-1). The first and last of these whistles were often slightly different to the centre ones, with the initial segment of the first whistle sometimes missing, and the last whistle in a series having a longer duration.

The estimate of the distance measure standard deviations for a class containing one whistle must be made based on the data in Table 8-1. Two hypotheses present themselves — either the standard deviations are proportional to the mean distance measures, or they are independent of the means. Using chi-squared analysis of the two hypotheses, $\chi^2 = 1743$ if the frequency distance measure is based on the mean, and 75.9 if it is independent. Although this figure is still very high, and by no means verifies the independence hypothesis, it is clearly more probable than the dependency one. These values could not be associated with probabilities, since the number of

degrees of freedom was meaningless. The values were calculated solely to identify the more likely of the two hypotheses.
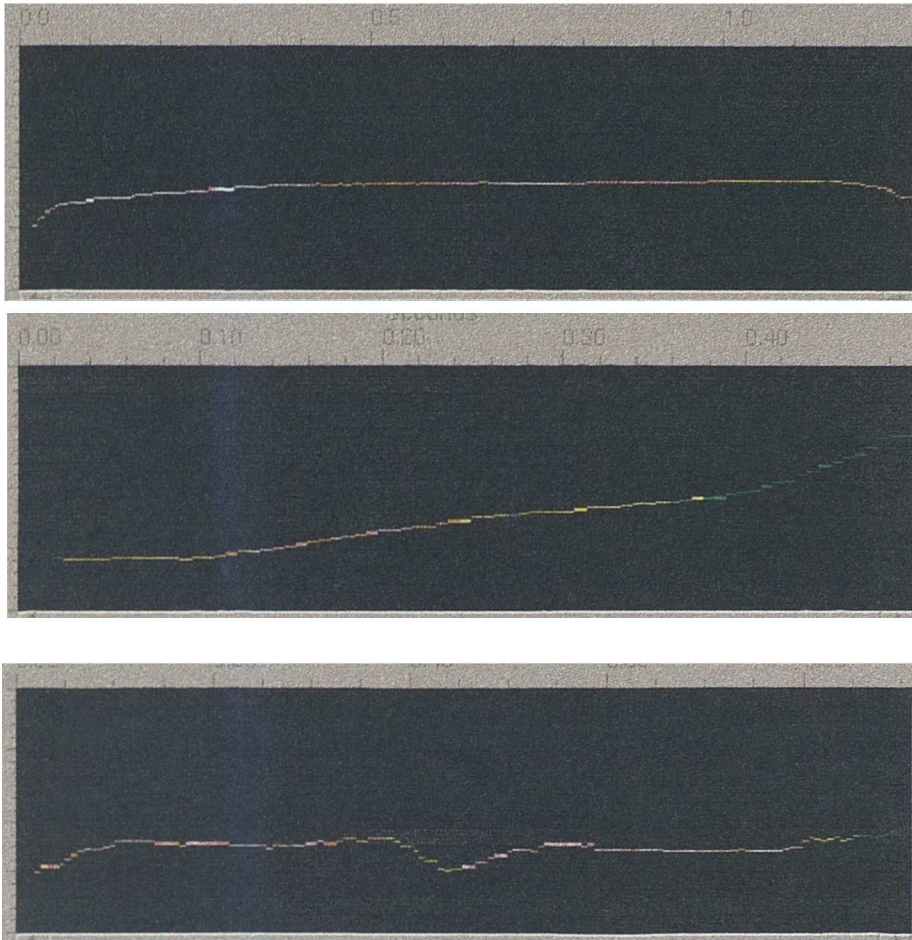

Figure 8-1: Typical whistle contours for dolphins Vindi, Nephele, and Sisu (top to bottom)

Similar calculations for the slope distance measure indicate a chi-squared of 212.7 for the dependency hypothesis, and 537.6 the independency. For the slope rate of chance distance measure, the chi-squared values are 8794 for dependency and 7643 for independency. From these results, it was determined to use initial estimates for standard deviations of the frequency measure and rate of change of slope measure which were independent of the calculated means. Values were chosen of 339.3 and 21739 respectively, and to set the initial estimate of the standard deviation for the frequency slope measure to 100.5% of the mean measure value.

Discussion

The figures for the standard deviations are all high compared to the means, which is unsurprising considering the measures are calculated from comparisons of pairs of

rising, flat, and falling segments. The standard deviations for each measure for the flat segments in a contour might be different to those taken from rising or falling segments, or the position of the segment within the whistle might have a bearing on the standard deviation. Additionally, whistles were included in classes to which they would not ordinarily have been placed based on their shape and structure, thus producing more upward pressure on the standard deviation values.

Assessment of Common Dolphin Groups

Recordings of groups of common dolphin (*Delphinus delphis*) were made during a related project named CETASEL, which is investigating the interaction of small cetaceans with trawl nets. Field trials of a tracking system in use around a trawl included a monitoring hydrophone attached in front of a trawl to monitor frequencies below around 30 kHz. Other hydrophones were attached to the trawl itself, but were modified to detect echolocation clicks.

Equipment and Data

Vocalisations were recorded on an RDAT recorder (Sony TCD-D7) with a 22 kHz bandwidth. Signals from the trawl-mounted hydrophone were preamplified (with a Benthos AQ4/AD743) before being sent via a coaxial cable some 450 metres back to the ship. Observers aboard the ship made records of encounters, which included details of sighting times, species, number of animals, sighting position relative to the ship, and direction of travel.

Whistles were noted between 15:06 and 15:17 on 10th October 1996 in the acoustic log, and comparison with the logs from the visual observers indicated three groups of common dolphin during that time, although not all were within range simultaneously. The vocalisations on the recordings seemed to fall into three periods of activity separated by intervals of quiet or only distant sounds from the dolphins. A human listener had previously detected the whistles on the tape, and these whistles were subsequently sampled onto computer. The three periods were termed A, B, and C, and contained 7, 12, and 30 extractable whistles respectively. The signal to noise of some whistles on the recordings was too low for satisfactory extraction, and so these were omitted from the totals. All other whistles were classified according to the
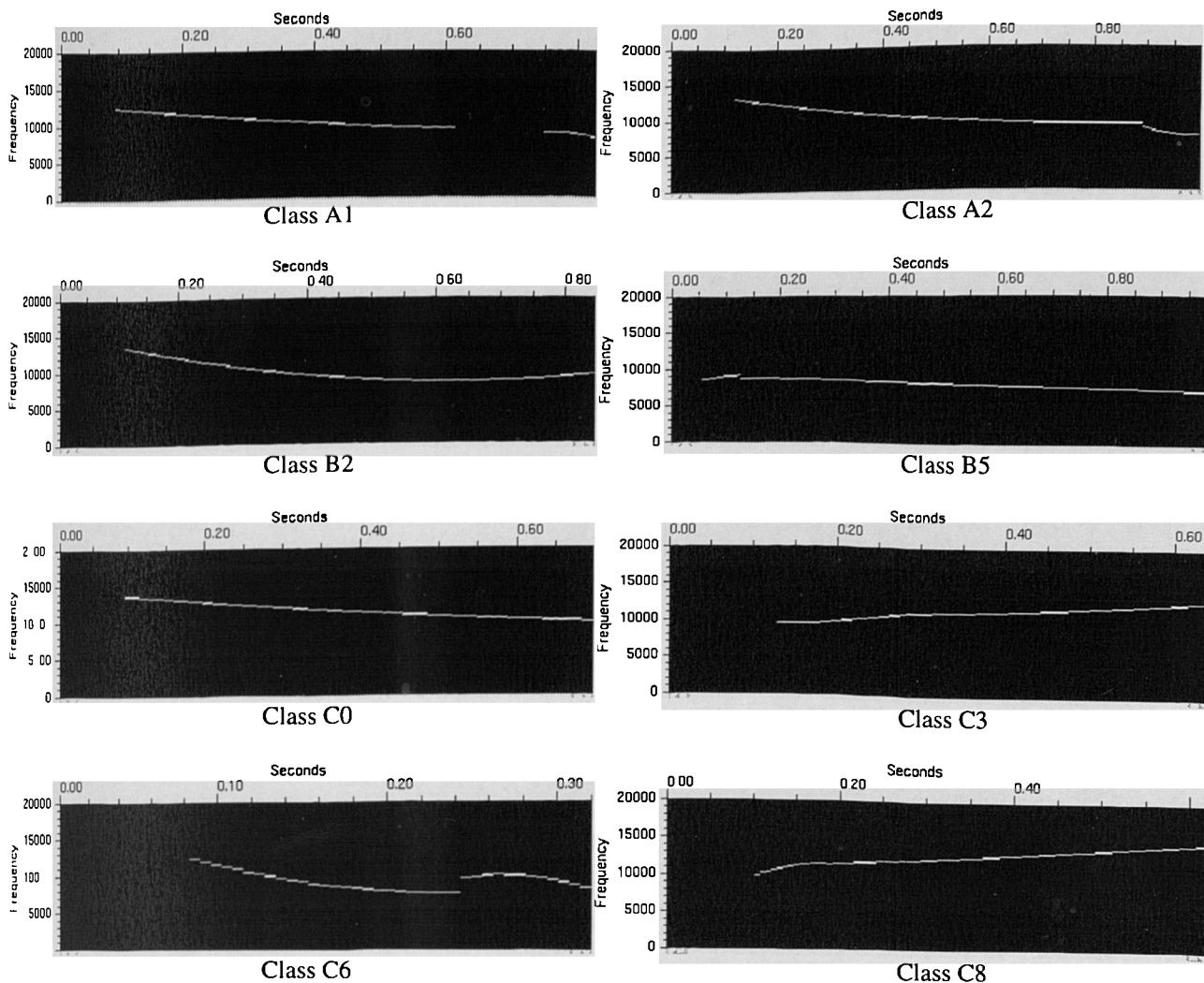
Figure 8-2: Typical whistle contours for each of the eight major classes.

automated classification procedure for each period, explained below. From these groups, the whistles that could not be isolated were three whistles from period B and two from period C, a failure rate just under 10%.

Whistle Group Classes

The whistles from each period were termed whistle 'groups', such that group A contained all whistles from period A, etc. The classification process described in the 'Whistle Classification' chapter (chapter 7) under 'Classification Decision' was applied to each group of whistles. Thus, the whistles within each group were assigned to one of a number of different classes. These classes and membership are shown graphically in Figure 8-2 and 8-3.
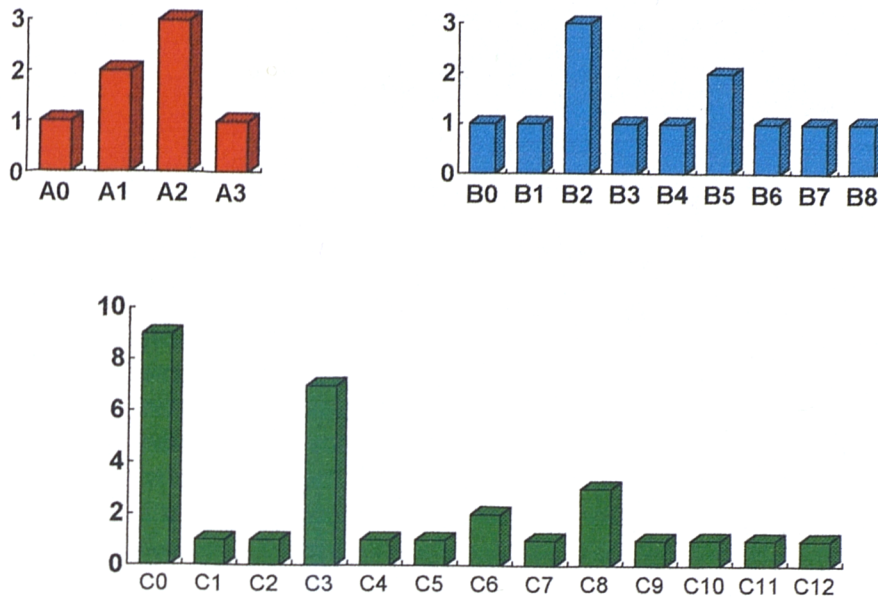
Figure 8-3: Classes and membership frequencies for the three dolphin groups.

Each class is labelled with the group letter and the class number within that group (starting at zero). Several classes within each group contained just one whistle, with a minority containing more than this. It is possible that some of these single member classes were due to 'aberrant' whistle from dolphins, i.e. a whistle type other than that individual's 'signature' whistle. Another hypothesis is that the whistles of this dolphin species are variable and not unique to individuals.

Cross-Group Comparison

The classes formed for each group were then cross-matched with those of the other groups. This was achieved by attempting to classify contours from one group into the classes already characterised for another. The results are shown in Table 8-2 for the three pairs of groups. A chi-squared analysis on the groups was made based on the null hypothesis that there was no difference in whistle distribution between classes from different groups.

	A0	A1	A2	A3	B0	B1	B2	B3	B4	B5	B6	B7	B8
Group A	1	2	3	1	0	0	0	0	(2)	0	0	0	0
Group B	0	0	0	0	1	1	3	1	1	2	1	1	1

$$\chi^2 = 16.9,\ 12\ \text{d.f.},\ p = 15\%$$

without (2) misclassified, $\chi^2 = 19.0$, p = 8.9%

	A0	A1	A2	A3	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Group A	1	2	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Group C	0	2	1	0	9	1	1	7	1	1	2	1	3	1	1	1	1

$$\chi^2 = 27.9,\ 16\ \text{d.f.},\ p = 3\%$$

	B0	B1	B2	B3	B4	B5	B6	B7	B8	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Group B	1	1	3	1	1	2	1	1	1	3	0	1	0	0	0	0	0	2	0	0	0	0
Group C	0	0	9	1	0	3	1	0	0	9	1	1	7	1	1	2	1	3	1	1	1	1

$$\chi^2 = 21.0,\ 21\ \text{d.f.},\ p = 44.5\%$$

Table 8-2: Cross-group matching of classes for the three whistle groups.

Chi-squared analysis reveals that it is unlikely that group A has the same whistle type distribution as either of the other two groups. None of the whistles from group B matched any of the classes from group A, although the software indicated that two whistles from group A fell into class B4. On further investigation, the software had misclassified these two whistles, since both whistles from group A had a single segment with a rising contour which had been characterised as 'flat', which matched the 'flat' of single falling contours from group B in class B4. If these two mismatches were removed from the table, the probability of the two group having the same whistle class distribution fell to 8.9%. Although not statistically significant due to the small sample sizes, this strongly suggests that groups A and B have different whistle characteristics.

Although three whistles out of 30 from group C were matched by classes formed from group A whistles, none of those from group A fell into group C classes. Chi-squared analysis of this table indicated a significant difference at the 5% level in the whistles from the two groups.

Similar analysis of classes and whistles from groups B and C produced a very ambiguous result. Further analysis was conducted on just those classes that contained more than one whistle, thus eliminating classes consisting of aberrant whistles, these being termed the 'major' classes for a group. The probability that whistles were evenly distributed across each group pair was recalculated with this new data in Table 8-3.

	A1	A2	B3	B4	B8
Group A	2	3	0	(2)	0
Group B	0	0	2	2	2

$\chi^2 = 8.98$, 4 d.f., p = 6.1%, (2.7% without (2))

	A1	A2	C0	C2	C3	C6	C8
Group A	2	3	0	0	0	0	0
Group C	2	1	9	1	7	2	3

$\chi^2 = 17.4$, 6 d.f., p = 0.8%

	B2	B3	B5	B6	C0	C2	C3	C6	C8
Group B	3	1	2	1	3	1	0	0	2
Group C	9	1	3	1	9	1	7	2	3

$\chi^2 = 5.91$, 8 d.f., p = 65.8%

Table 8-3: Comparison of 'major' classes between the three whistle groups.

Classes B2 and C0, B3 and C2, and B5 and C8 all contained the same whistles, suggesting that these class pairs were identical, although classes C3 and C6 contained no whistles from the other group. Excluding C3 and C6, the probability that the two groups contained the same whistle type distribution was 94.4%. One explanation is that the two groups of dolphins were recorded simultaneously for a time, and whistles from C3 and C6 belonged solely to the second of the two groups.

One last conclusion was that the common dolphin uses whistles that are group specific, and possibly specific to individuals within that group. This suggests that this species might be added to the list of those small cetaceans that employ signature whistles. This area should receive further research, since little is known for this oceanic species.

Class no.	No. of whistles	Encoding sequence	Whistle id's
0	5	'UD- -'	mf5w10a, mf5w13a, mf5w14, mf5w16b, mf5w17
1	1	'-UD'	mf5w10b
2	1	'U-D- D-'	mf5w11a
3	1	'U-D-U D-U'	mf5w11b
4	2	'UD- -U'	mf5w12, mf5w16a
5	1	'UDU -'	mf5w13b
6	1	'U-D- -U'	mf5w15
7	9	'-U'	mf5w18a, mf5w18b, mf5w18c, mf7w16, mf7w23, mf8w01, mf8w02, mf8w03, mf8w16
8	22	'U'	mf5w28d, mf5w28e, mf5w28f, mf5w44, mf7w50b, mf8w05a, mf8w15b, mf8w15c, mf8w23b, mf8w23d, mf8w24a
		'U-U'	mf5w19a, mf7w29, mf7w36, mf7w73
		'U-'	mf5w29, mf7w20, mf7w37a, mf7w37b, mf7w46, mf7w69c, mf8w06
9	1	'U-U- -'	mf5w19b
10	1	'-U DU'	mf5w19c
11	1	'U-D U-D U-'	mf5w20a
12	1	'UDU- - U- - -'	mf5w20b
13	1	'U-U'	mf5w21a
14	1	'-U-'	mf5w21b
15	1	'U- - U-D U- -'	mf5w23
16	1	'U - - -'	mf5w24
17	1	'- - - - D - -'	mf5w25
18	1	'- - - - U'	mf5w26
19	1	'-U-U'	mf5w27
20	1	'-D U'	mf5w28a
21	1	'- U'	mf5w28b
22	1	'U- - U'	mf5w28c
23	1	'D-U- -U'	mf5w30
24	6	'-'	mf5w31, mf5w36, mf7w38, mf7w40, mf7w41, mf7w47a
25	5	'DU'	mf5w34, mf5w45, mf7w45, mf7w48a, mf8w15a
26	1	'DU'	mf5w35
27	1	'U-D'	mf5w43
28	1	'D-U'	mf7w01
29	1	'D-'	mf7w02
30	1	'-D U-U U'	mf7w04
31	1	'DU DU'	mf7w09
32	1	'U U'	mf7w22a
33	1	'-U- -'	mf7w22b
34	1	' U-D D'	mf7w22c
35	8	'U-D'	mf7w47b, mf7w48b, mf7w50a, mf7w54, mf7w59, mf7w67a, mf7w67b, mf8w05b
36	1	'DU-'	mf7w55
37	1	'UDU-D-'	mf7w57
38	1	'U-DU -'	mf7w61
39	1	'U- D'	mf7w69a
40	1	'U- D'	mf7w69b
41	1	'U -'	mf7w69d
42	1	'D-U-'	mf7w74
43	1	'D U DU DU DU'	mf8w08
44	1	'U-DU-D'	mf8w10
45	1	'DU-U-U D'	mf8w11
46	1	'-D'	mf8w19
47	1	'D-D'	mf8w20
48	1	'U U'	mf8w23a
49	1	'DUDU-'	mf8w23c
50	1	'- -U'	mf8w24b

Table 8-4: Moray Firth 1992, tapes 5, 7 & 8 whistle classes. See table 6-1 for key.

Bottlenose Dolphins in the Moray Firth

Whistles recorded on several tape recordings made during the field trials in the Moray Firth during 1992 were sampled and catalogued by Ilka Giannikos (a visiting

researcher from Germany), and whistles from three of these tapes were used as input to the classification routines. Recordings of the dolphins' vocalisations were made on a four-track Racal Store 4DS instrumentation recorder at a tape speed of 7½ inches per second, and included an EBU time-code with a 1/25th second resolution. A number of sources were used, including hydrophone radio telemetry links (sonobuoys) deployed approximately 500 metres away from the land-based observation point, and from directly cabled hydrophones. Contributions from sea-state noise were reduced by filtering at source with a high-pass filter with a 3 dB point at 1 kHz, and the signals were corrected to provide a nominally flat frequency response from this value to at least 20 kHz. When dolphins were present, the time, group compositions (number of adults and juveniles), and apparent behaviour were logged.

Out of 121 recorded samples, 101 whistles could be satisfactorily extracted with the analysis software. There were samples where it was apparent to the ear that there were whistles present, but they were very distant and lost in background noise. Few of these could be isolated using the filtering and contour extraction routines, although this proved possible with some samples.

Classification of the resulting encoded whistle contours in a similar manner to the previous whistles from common dolphins resulted in whistles being classified as 'new', even though the associated probability was lower than 50%. This situation would often occur where the whistle was matched reasonably well by two classes, frequently in the range 25–35%, and a 'new' class probability of between 35% and 40%. For the previous classification protocol, the whistle would be placed in a new class, even though the probability for this is quite low. Since the combined probability that the whistle type had previously been observed was higher (albeit split over two or more classes), the protocol was modified such that the whistle was placed in the post probable existing class when the probability of it belonging to a 'new' class fell below 50%.

Classification Results

Whistles were labelled with the recording location (mf, for Moray Firth), tape number, and sample number, e.g. mf5w31. Where more than one whistle was contained within one sample, these were distinguished by a letter suffix (e.g. mf5w31a, mf5w31b).

Whistle types were divided into 51 classes for 101 whistle contours (Table 8-4), but only 7 of these (classes 0, 4, 7, 8, 24, 25, and 35) formed 'major' classes as defined for the common dolphin classes above.
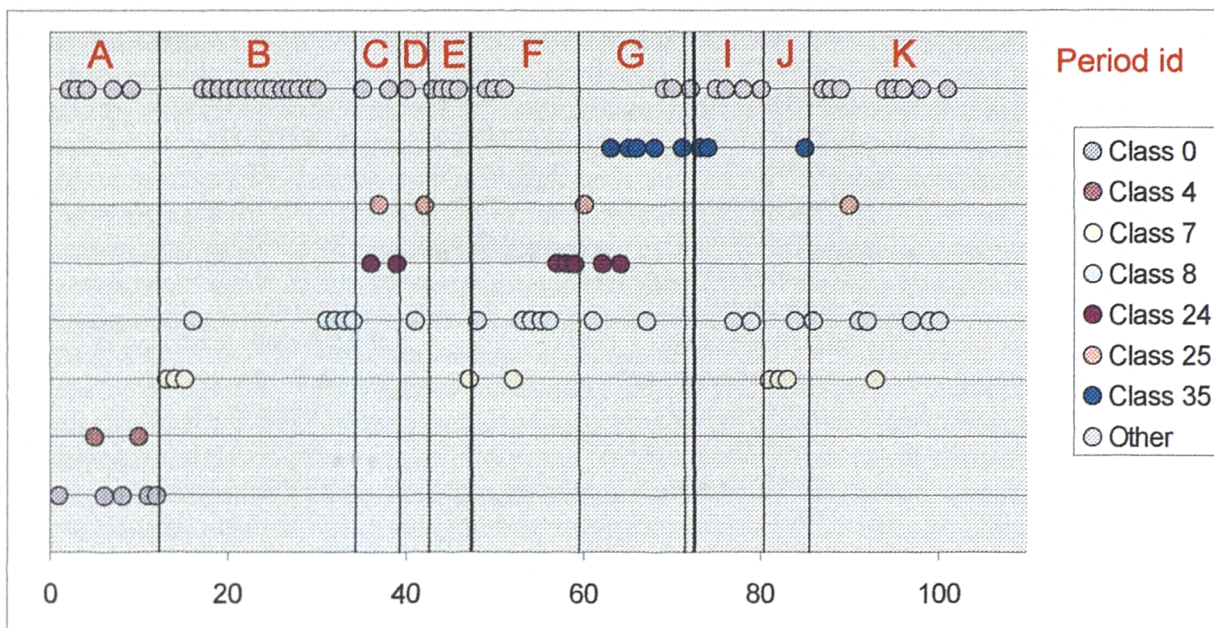

Figure 8-4: Occurrence times of the major classes (fine lines separate recording periods, thick lines days).

Plotting the class occurrence on a timeline for the major classes indicates those classes that might be useful for distinguishing between groups of dolphins (Figure 8-4). The thick vertical bars separate the three days of recordings that the tapes represent (14th 16th September 1992), and the finer lines separate the recording periods when dolphins were present. Table 8-5 lists the times for each recording period.

Date	Recording period	Whistle range
14 9	A: 12:40–12:41	Mf5w10–17
	B: 12:50–12:52	Mf5w18–29
	C: 12:55–13:07	Mf5w30–36
	D: 13:29-13:31	Mf5w43–45
	E: 20:26–20:29	Mf7w01–16
15 9	F: 08:45–08:52	mf7w20–41
	G: 09:32–09:35	Mf7w45–59
	H: 18:20	Mf7w61
16/9	I: 14:28–14:31	Mf7w67–74
	J: 14:49–14:52	Mf8w01–05
	K: 14:57 15:09	Mf8w06–24

Table 8-5: Recording times for the Moray Firth whistle

Construction of a chi-squared table for the different periods (Table 8-6) did not reveal any strong conclusions about group identities between recording times. There were four values in the table that produced statistically significant results. Periods A and B (labelled alphabetically from the start of the recordings) appear to contain significantly different whistle classes, and similarly for periods B and G. Periods D and E and periods D and J would appear to be very similar in whistle types, but the small class totals for those periods have a disproportionate effect on the result when Yates' correction for integer sample sizes is taken into account. Uncorrected values for periods D and E would be just 20.5%, and for D and J 40.4%.

	B	C	D	E	F	G	H	I	J	K
A	3.8	51.4	81.5	67.8	13.8	20.6	64.2	39.1	26.5	13.8
B		26.5	66.8	84.1	16.7	4.8	46.9	44.6	17.0	93.0
C			91.2	88.1	85.8	75.4	63.8	82.4	80.1	39.7
D				95.6	84.8	91.9	71.6	92.8	95.7	75.7
E					45.2	58.3	32.7	92.8	47.5	51.8
F						47.7	38.7	72.0	58.5	58.3
G							33.3	88.9	44.9	27.0
H								67.4	49.3	6.9
I									25.8	75.0
J										24.5

Table 8-6: Chi-squared values (Yates correction applied) for percentage similarity between periods (shown in 8-3). Significant values are highlighted.

It might be possible to improve these values if the 'Other' class were removed from the calculations. This would also have the effect of considerably reducing the number of whistles within each period and thus make any probability values calculated even more tentative.

Discussion of the Moray Firth Results

Comparison of the times for the recordings and the class occurrence diagram in Figure 8-4 suggests that class 8 is present in all periods except A, whereas classes 0 and 4 seem to be specific to period A. Period B seems to begin with classes 7 and 8, and then a large number of aberrant whistles occur before these two classes are displaced by three whistles of classes 24 and 25 in period C, just 3 minutes later. Another whistle from class 25 was heard around 13:30, some 23 minutes later, although with only three whistles in this period it is difficult to assign any significance to this occurrence.

On the following day, classes 24 and 25 can again be seen occurring towards 08:52 and then again at 09:32. Class 35 then replaces these classes towards the end of period G, and this new class reappears the following day on three occasions — 14:28 (twice) and 14:52.

Although the number of whistles that fall into major classes within each period is too small to produce statistically significant results, the class figure (Figure 8-4) strongly suggests the presence of three groups of dolphins:

Group 1: Period A is characterised by whistles from classes 0 and 4, and a lack of whistles from class 8, which can otherwise be seen spread throughout the different periods. This dolphin group was seen only once between 12:40 and 12:41 on 14/9/92.

Group 2: Period C is characterised by whistles from classes 24 and 25, and this group passed through the recording area between 12:55 and 13:07 on 14/9/92, and possibly between 13:29 and 13:31 the same afternoon. It reappeared again on 15/9/92 towards the end of the recording period from 08:45 to 08:52, and then again at the start of the period 09:32 to 09:35 before being replaced by Group 3.

Group 3: This group replaced Group 2 during the 09:32 to 09:35 recording period on 15/9/92, and contained whistles in class 35. It reappeared the following day (16/9/92) at 14:28 to 14:312, and possibly again from 14:49 to 14:52.

Discussion

The restriction on the segment type to being 'up', 'down', or 'flat' proved to have a negative effect on the classification. There were instances when parts of the whistle contour had a slope close to that used to separate 'up' or 'down' from 'flat', and this could result in the inclusion of one or two extra segments in the resultant encoded contour (e.g. 'up' → 'up' 'flat' 'up'). These extra segments could have a large effect on the probability calculated by the hidden Markov model routine. If a larger number of slope-based segment types were provided, this would allow a more flexible representation of the whistle. Alternatively, segments could be defined by areas of similar slope, and the slope within each attached as an attribute to be used within the HMM routine.

In many forms of classification, both automatic and manual, there is a tendency for classifiers either to place similar objects together in the same class or to split them into separate classes. Inspection of the table showing the classes for the common dolphin data (Table 8-2) shows that the method employed here tends to place whistles into different classes based almost entirely on their segment sequence, suggesting that this method falls into the 'splitter' variety of classifiers. The effect of this splitting could be seen with the bottlenose dolphin whistles from the 1992 Moray Firth trial. Even though the chance that a whistle belonged to a new class was less than 50%, the probability of individual class membership fell below this since the probability was spread over a number of 'split' classes.

Several whistles which one might have classed together were instead 'split' by the program into separate classes, since some contained extra segments due to the slope of parts of their contour lying close to the boundary between the 'flat' and 'up' or 'down' labels. It would be possible to weight the HMM transitions to mitigate this effect, by providing a probability for alternative transitions based on the segment slope. Thus, an 'up' segment with a slope close to the boundary between 'up' and 'flat' would have a probability close to 50% of having all or part of that segment labelled 'flat'. Some empirical studies would be required to determine the variability of such segments.

The three distance measures that are calculated by comparing whistle segments, are converted into probabilities based on standard deviations calculated over all the segments within a whistle. However, this approach assumes that the standard deviations for the three measures are equal over all segments within the whistle, which in practice is rarely the case. Storing the standard deviations for each segment in a whistle would seem to be the obvious solution, but when the segment sequence may change within one class to accommodate differences between whistles, it is unclear what sequence of standard deviations should be used for a corresponding segment sequence.

Summary

Standard deviations were required for the distance measures used to calculate the similarity of two matching segments. These were calculated from recordings made of

whistles from three known dolphins, where the whistle contours were extracted, encoded into segments, and then forced into classes for each dolphin based on observations carried out at the time of the recordings. The differences between segments could then be quantified and the standard deviations calculated.

The whistles from three groups of common dolphins recorded in the North Atlantic were analysed. The classification of the whistles was made based on the class that gave the highest membership probability, or the whistle was placed into a new class if this was more probable. The results showed that the first of the three groups had significantly different whistles from those of the other two, but that the whistles of the second and third groups were not significantly different. Further analysis showed that whistles were recorded that were significantly different between the groups, but also that there were whistles from the first group that were recorded in the second. The suggested explanation for this is that both groups were within range of the hydrophone for a period of time, and the whistles from this overlapping interval were assigned to the third group.

Since the whistles from different groups of common dolphins could be differentiated, this suggests that there is at least some group coherence to whistle production in this species, and possibly that they use signature whistles in a similar way to the bottlenose dolphin.

Recordings from bottlenose dolphins were also analysed which were made during field trials carried out in the Moray Firth in 1992. 101 whistles were satisfactorily extracted from tapes made on three consecutive days, and classifications made. The majority of these whistles fell into classes that consisted on one whistle, but seven classes consisted of more than one whistle and were termed 'major' classes. Although only two recording periods showed any statistical significance in their whistles from the rest of the periods, plotting the whistle classes against time suggested three identifiable groups. One group occurred on one day only, whilst the other two were observed on consecutive days.

The main drawback with the method of classification as it stands is that whistles are too often placed in different classes when a human classifier would place them in the same one. If this could be rectified, it might be possible to gain more statistically significant results on inter-group similarities or dissimilarities. This might

be achieved by using more segment types that reflect smaller slope ranges, and combining this with a modification to the hidden Markov model that initially allows whistles with similar segments (in slope) to also have a high probability.

Another more minor discrepancy occurs when a segment's slope is close to a type boundary, where small deviations can change the segment sequence. This can again be solved by allowing a range of values around the average slope to take on a range of probabilities.

Chapter 9: Conclusions

Introduction

The *raison d'être* behind this research was to provide a novel research tool which would aid in the study of changes in behaviour over time of dolphins in the wild. No such acoustic tool previously existed for the quantitative measurement of differences in whistle structure. During the research, however, Reiss and McCowan proposed a different method for whistle analysis in the United States. This technique involved sampling the whistle's frequency at twenty points equidistant along the contour, thus normalising over time. Due to the high computational overhead of the Reiss and McCowan method, it is believed that the method presented here allows for a more complete representation of the whistle, as well as it allowing comparison of larger databases of whistles.

Achievement of the Research Objectives

An automated technique has been developed that enables the computer to monitor a recording and indicate those times where tonal signals, such as whistles, are present. The recording can then be replayed to allow sampling to take place of the relevant areas of the tape. This process can take place at the full replay speed of the recording, rather than making use of a human listener to identify whistles. In the case of a human listener, the recording would have to be replayed at half speed in order that all whistles would fall inside the human hearing range.

An 'inertially' driven routine has been devised for extracting the frequency-time-intensity contour of the whistle, and digital filters for background noise reduction have been developed to reduce these effects on the extraction process. This method allows simultaneous whistles to be separately extracted under some conditions, even when the contours cross each other in frequency, provided that the frequency slopes are distinct at such times.

An encoding system based on a whistle's shape has enabled whistles to be represented in a compact form, suitable for storage in a database to allow searching

for comparison with candidate whistles. Both the overall shape as well as a more detailed description are available, allowing either a quick comparison of whistles (for instance, by ordering the database by segment sequences), or a more detailed analysis of similarities and differences (using the stored curves for each segment).

Whistle classes have been constructed based on syntactic representations based on hidden Markov models of segment sequences, which are combined with a statistical approach, based on distance measures between segments within the whistles. The probability can then be calculated for a whistle chosen at random from each class having a match to a candidate whistle. Empirical analysis also has enabled an estimate of the probability for a class not represented in the database, and this final value enables the probability that the whistle is a member of each class to be calculated.

A number of whistles could then be combined together into groups based on times at which they were recorded, and the differences in whistle types between groups could be assessed for significance using their probability of occurrence. This was achieved by constructing classes based upon one group, and then attempting classification of whistles from another group with those pre-existing classes. In this way a table of expected whistle-type occurrences could be constructed, and assessed for significance.

The primary aim of determining the difference between groups of dolphins has therefore been achieved. The significance level can be set at whatever the researcher desires, (for example: greater than 90% indicates similarity; less than 10%, difference; other values being undetermined in outcome).

Practical Results

The routines presented as part of this thesis used data from dolphins recorded in dolphinariums in their development. These signals are easier to analyse for contour extraction than those of wild dolphins since the signal to noise ratio is normally much greater. The classification process was made easier, too, by knowing the number of vocalising animals.

Attempting analysis of data from the wild allowed the techniques developed in this way to be assessed. Two trials were carried out, using sets of recordings of

common dolphins from a moving ship in the north Atlantic, and secondly bottlenose dolphins from the Moray Firth, Scotland, recorded from a static position.

A total of 49 whistles were extracted from the common dolphin recordings, and split into three groups separated by long periods of silence. The whistles were encoded, and classes were formed and stored for each of the three groups. The probability was then calculated that whistles from the other groups fell into those classes, and vice versa, forming occurrence tables of whistle types between groups. Statistical analysis was carried out on the tables, revealing that one group contained significantly different whistle to either of the others, but no similarity/dissimilarity decision between the remaining two groups. Further analysis revealed that the latter two groups did contain distinct whistle types, but that there had been a period where both groups appeared to have been recorded at the same time.

101 whistles from bottlenose dolphins were extracted from tapes made over three days in the Moray Firth. The most frequent classes were plotted on a time line, and any patterns assessed. It was difficult to test the significance of these results since the data contained relatively few whistles for each group. Three groups were, however, identified due to distinctive whistle types. Group 'A' was found to occur once only on the first of the three days, group 'C' occurred on one afternoon, and then twice again the following morning (day two), and group 'G' followed group 'C' on the second day, and then reappeared on the afternoon of the third day.

Once analysis and classification have been completed, it has been shown that acoustic recordings of wild dolphins can be used for identification of groups from one occurrence to the next. This can also be applied to the common dolphin, a species that has not been reported widely as making use of signature whistles (or other identifying sounds). It seems probable that the vocalisations of many other small odontocetes might be amenable to similar treatment, greatly extended the scope of this software.

Acoustic Or Visual Means of Identification

The traditional method of field identification of individual cetaceans is using photo identification. For behavioural research, however, its main drawbacks are:

- Close proximity of the photographer to the target animal

- Need for a stable platform

- *An experienced photographer is needed in order not to miss an often brief surfacing*

- *If at sea, the presence of a boat can disrupt the target animals' natural behaviour*

Acoustic recognition overcomes many of these problems, provided one has a method of identifying the target animal,. The receiving hydrophone need not be close to the recording equipment, and can be linked via cable or using a radio transmitter. The hydrophone can be attached to a buoy, and the platform for the recording equipment need not be as stable as for photographic identification. Recording can begin whenever whistles are heard over the monitor speakers, and therefore this obviates the risk of missing potential identification opportunities as might happen with a photographer.

Control studies of bottlenose dolphins swimming through the study area in the Moray Firth, Scotland showed that the dolphins had no reaction to the presence of a sonobuoy in their vicinity. Even at sea, a sonobuoy can be towed several hundred metres behind a boat and thus reduce the disturbance to the dolphins' natural behaviour.

There are also disadvantages to the use of an acoustic system. Since in general one dolphin does not always produce the same whistle, the confidence in any identification is not as high as for photographic identification. When identification of a group of animals is required, this problem can be overcome by examining a number of whistles from different groups. In this way, the increased number of whistles available for comparison can be used with a statistical significance test to quantify the effect of similarities or differences in the numbers of different types of whistles.

In summary, if the identity of individual dolphins is required, acoustic identification techniques alone are insufficient due to the variable nature of the dolphins' whistles, and possibility of distinctive whistles being used by more than one individual. Where the identity of groups (rather than individuals) is required, and the behaviour of the animals is also part of the study, acoustic identification can be a

much better option. This especially applies when there is no available observation point close enough to the animals for satisfactory photography without risk of affecting their behaviour.

Proposals for Future Work

Although the current software package allows the extraction and analysis of the majority of whistles recorded in the field, there have been several instances where it has been possible to visually determine the outline of a contour, but where the software has been unable to extract it. Sometimes this has been due to the signal to noise ratio of the whistle being too faint for the software to detect, combined with a contour that modulates its frequency too fast for the spectral exponential averages to react.

One solution to this problem is to calculate such correlations not only for stationary frequency signals, but also for signals that change at rates within those typical for small cetaceans. This will increase the amount of required computational effort during the detection process, but if limited to the off-line section of the analysis should still be within acceptable limits.

The spectrogram itself is calculated on a consecutive series of discrete windows through the data. If these were allowed to overlap to a degree, so some of the data from the previous time partition is still present for the next calculation, then the time resolution of the spectrogram will be increased with no consequent loss in spectral resolution. The payment for this falls again in terms of increased computational effort, but since computer speed appears to be increasing at a steady rate, this factor might shortly cease to be a serious drawback.

It is believed that the inertial model for whistle contour following performs the task very well. When overlapping contours occur, however, sometimes the choice for the initial starting point falls on a different contour to the one that is desired. If a system were to be used whereby after a whistle contour had been isolated, then that contour's intensities could be removed from a spectrogram (or set back to some background noise level), then this would enable further contours to be extracted from the same signal. Much of the data from the field trials produced either single-whistle

recordings, or recordings of multiple whistles where one whistle predominated, so this technique was not developed further.

The system of classifying the whistle segments into three types dependent on the direction of travel, appeared sometimes to weight the probabilities too heavily, especially when segment occurred close to the boundary between types. A system whereby a larger selection of segment types is available would help to obviate this problem. The weightings of the hidden Markov model could be initialised to accommodate segments of similar slope until a sufficient number of members had been assigned to it. Alternatively, a continuous hidden Markov model could be used instead of the discrete type currently in use, and the average slope of each segment used as the input to the HMM. The new problem of the positioning of segment boundaries would then need to be addressed.

The whistle classification package has now reached a stage where it has become useful for determining dolphin group identities. Currently, only a small amount of the data collected from the Moray Firth trials has been analysed. It is believed that a great deal more could be learned from this data about the habituation effects of the acoustic reflectors under study. The software has also lent itself to shipboard trials, and has been successfully used with vocalisations from the common dolphin. It would seem reasonable to expect that a number of other small odontocetes that use whistles for communication might also be amenable to classification using these methods.

Acknowledgements

Appendix A - Echolocation Literature Review

Literature Available on the Echolocation Characteristics of the Harbour Porpoise (*Phocoena phocoena*) and the Common Dolphin (*Delphinus delphis*)

C.R. Sturtivant, S. Datta and A.D. Goodson

Signal Processing Research Group – Bioacoustics & Sonar, Electronic Engineering Department, Loughborough University of Technology, U.K.

Introduction

The design of passive acoustic deterrents to mitigate the problem of small cetacean by-catch in commercial gill-nets requires detailed knowledge of the echolocation characteristics of the species at risk. In U.K. waters the cetaceans perceived to be at greatest risk are the harbour porpoise (*Phocoena phocoena*), and the common dolphin (*Delphinus delphis*).

The primary purpose of echolocation in odontocetes would appear to be in the detection and pursuit of prey. The foraging habits of the two species differ markedly from each other, which thus might suggest differing echolocation adaptations may have evolved in response. *Delphinus* is a pelagic animal generally feeding nocturnally on a variety of small schooling fish and squid, including sardines, anchovies, herring and pilchards. *Phocoena*, in contrast, is usually found in shallow water bays, rivers, and estuaries, and frequently feeds near the sea bed on less gregarious prey, in a potentially more reverberant environment.

Although the echolocation capabilities of the bottlenose dolphin (*Tursiops truncatus*) have been well studied, less data is available for *Phocoena*, and data for *Delphinus* is even more sparse. Evans and Awbrey (1988) stated that other than establishing that the common dolphin is very vocal, little or no work on its echolocation and associated behaviour has been done. Indeed, much of the current information on *Delphinus* comes from research written in Russian and carried out in the former Soviet Union.

This paper presents the results of a recent literature search on the echolocation abilities of these two species, and points out some of the physical factors which might have shaped them. It was not possible to obtain all papers relevant to this area (especially those from the former Soviet Union), but where this was the case much of the data was found in published reviews which were a valuable secondary source.

Physical Factors Influencing Echolocation Ability

Amundin *et al.* (1988) reported that the porpoise's repertoire consists exclusively of clicks, and does not contain any of the whistles common to many other small odontocetes. In his study Wiersma (1988) noted that the small bandwidth compared to click duration of many cetacean echolocation signals resulted in the best possible

signal to noise ratio in broad background noise. Moreover, *Phocoena*'s signal has a relatively narrow bandwidth and long duration when compared even to other cetaceans, suggesting an echolocation system where the signal to noise ratio is more critical than most.

Due to the click's "monochromatic" nature, Evans (1973) suggested that the harbour porpoise's sonar is primarily used for *detection* of targets, due to insufficient bandwidth to enable classification to take place on spectral "coloration". The echo from a target will have a frequency dependent component affected by differences in the target's size, swim-bladder shape and volume, which *Phocoena*'s narrowband sonar will be unable to exploit.

Amundin (1991b) suggested that *Phocoena*'s specialised sonar and absence of FM signals has developed due to the needs of the coastal habitat outweighing the consequent restrictions in vocal communications which are a feature of more gregarious species of cetaceans. He also noted that large schools would not be favourable for the capture of *Phocoena*'s bottom-living prey, which Hohn (1990) indicated were small schooling fish such as herring and anchovy.

It seems probable that *Phocoena*, due to its small size, may be limited in the intensity of the sounds it is able to produce. If this limits the total available energy for the pulse, a narrow band echolocation signal (and matched receiver) would thus produce the signal to noise ratio required at a longer range than for a wider band signal. The high frequency component of the porpoise's click would provide good resolution of objects down to the order of one wavelength in size (11.5 mm at 130 kHz). This frequency region has also been noted as relatively quiet, being above sea state noise and below the start of the thermal noise region (reference?).

Delphinus was noted by Evans and Awbrey (1988) to be very vocal during the day, producing whistles, squawks, clicks, and squeals, but then moved predominantly to echolocation trains at night. They generally forage on squid and small schooling fish around 20 cm in length, including sardines, anchovies, herring and pilchards, and in some areas organisms from the deep scattering layer which have risen towards the surface at night (Jefferson and Leatherwood 1990). *Delphinus* would therefore be expected to search over relatively long ranges with its sonar during foraging activity. It was also expected that its greater body size would enable this dolphin to produce a more intense echolocation click than that of the harbour porpoise. In low reverberant conditions, a louder signal would allow an increased bandwidth for the same signal to noise ratio (providing additional spectral information on the target), or a longer detection range for the same bandwidth, or some combination of the two. Indeed, frequency spectra for *Delphinus* are wideband, and appear more like those of *Tursiops* than *Phocoena* (Wood and Evans 1980, Evans and Awbrey 1988).

Evans and Awbrey (1988) also reported that moving schools of *Delphinus* were heard to click synchronously, with ranks within a school producing click trains almost in unison. This would avoid the click from a neighbouring dolphin masking some long range echo, as the sensitivity of the echolocation receptor would be decreased for some time by such a loud sound.

To summarise: the harbour porpoise forages near the sea bed in a reverberant environment, may be limited in the intensity of its echolocation due to physical size, and uses a narrow band echolocation signal which provides increased detection range

at the expense of target classification ability; the common dolphin forages in less reverberant conditions than *Phocoena*, has a signal with a wider bandwidth, and should thus be able classify targets at longer range.

Auditory Sensitivity

Behavioural techniques using a "go/no-go" criteria were used by Andersen (1970) to record the minimum detectable sound intensities for *Phocoena*. The results showed responses to tonal sounds in the range 1-150 kHz, and have been reproduced in Figure 1. Wood and Evans (1980) also reported the work done by Sukhoruncheko (1973) who determined that *Phocoena* had maximum auditory sensitivity at 8 kHz, 32 kHz, and 64 kHz. Although evidence of discrete sensitivities for *Phocoena* has not been encountered in other papers, the data from Andersen does indicate 8 kHz and 32 kHz as the frequencies of maximum sensitivity. One possibility could be that discrete points of sensitivity found by Sukhoruchenko may be artefacts produced by harmonics in the sound generation system.

Voronov and Stosman (1986) (as referenced by Evans and Awbrey 1988) used *Phocoena*'s brain-stem response to determine its audiogram (Figure 2). The results show a response sharply attuned to the frequencies used in their narrow band echolocation click (in the range 120 kHz to 140 kHz), which they also found for the Indus river dolphin (*Platanista indi*) and the Amazon river dolphin (*Inia geoffrensis*). Popov *et al.* (1986) conducted similar evoked potential experiments with *Phocoena* which showed increases in threshold sensitivity of approximately 15 dB at both 30 kHz and at 125 kHz. They also reported that certain areas of the brain investigated showed an increase in evoked response with a decrease in sound duration, whereas other regions produced the more expected increase in response with increasing duration
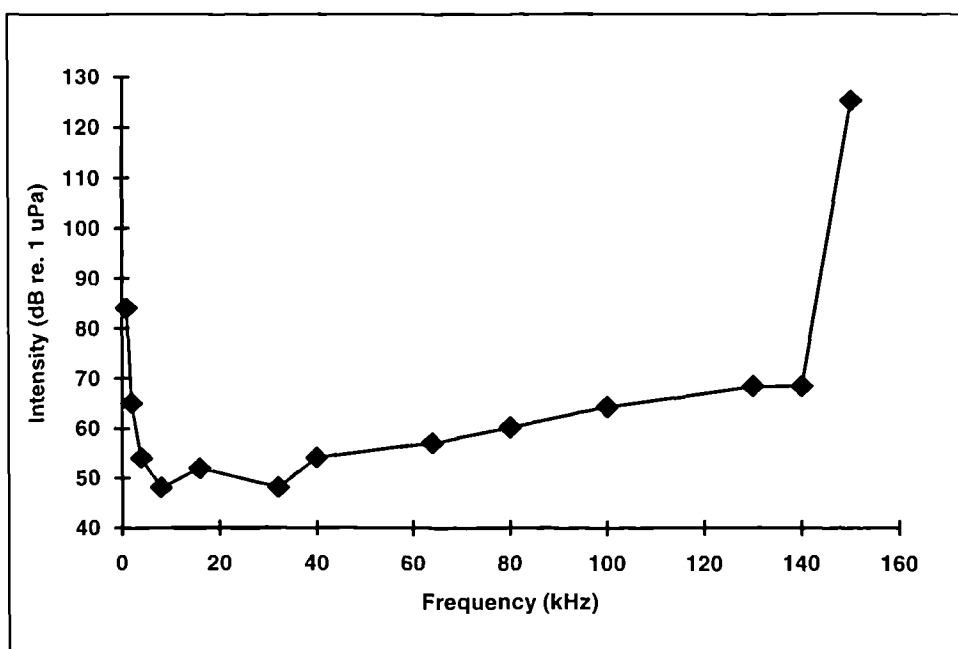
Figure 1
Phocoena phocoena *behavioural audiogram (taken from data in Andersen 1970)*

The brain-stem audiogram for this species is distinctly different from Andersens's behavioural audiogram in Figure 1. These differences can be explained if there exists at least two different mechanisms of sound detection — the one measured by Voronov and Stosman being tuned for reception of echolocation clicks, and Andersen's measurements being from the animal's response to all its receptors. The existence of a low-frequency receptor for *Tursiops* was recently proposed by Turl (1993), which the dolphin apparently employed in an auditory context.
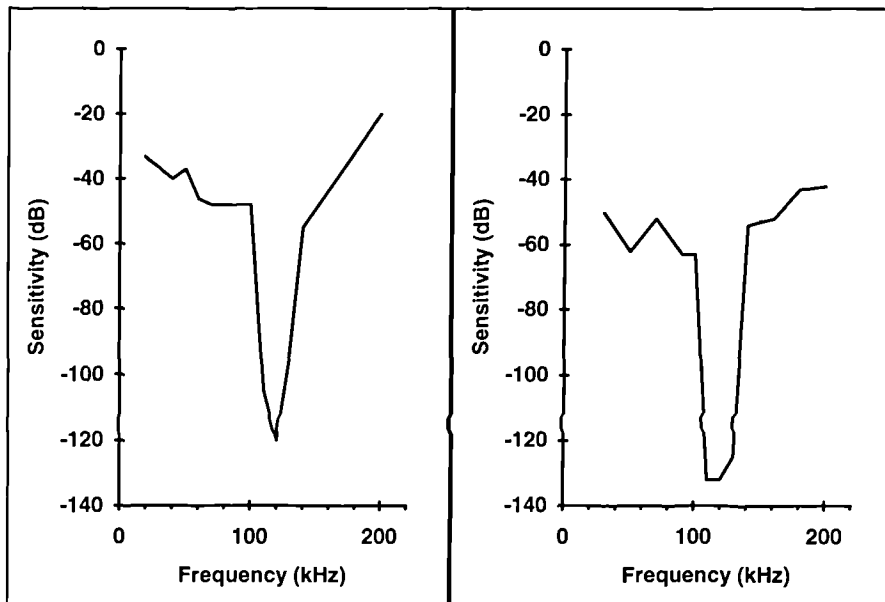
Figure 2

Frequency thresholds for two Azov Sea harbour porpoises (Phocoena phocoena) taken from data in Evans 1988 (Voronov & Stosman 1986)

Delphinus's hearing range was measured by Bel'kovich and Solntseva (1970) (reported by Wood and Evans 1980). They found a response to frequencies between 100 Hz to 280 kHz, with maximum sensitivity in the 60 kHz to 100 kHz region, although Wood and Evans did not specify whether this was a brain-stem or behavioural audiogram. If the sensitivities at the higher frequencies (up to 280 kHz) are significant, this would appear to give *Delphinus* a very wide hearing range when compared to other species of small cetacean for which an audiogram has been determined.

Echolocation Click Production Mechanism and Beam Pattern

Phocoena's echolocation click contains energy in two separate frequency ranges, noted by several authors (Amundin 1991b; Kamminga 1988). The low frequency component has been recorded at around 2 kHz, and the high frequency component at 120-140 kHz, although Kamminga and Wiersma (1981) recorded one individual with a 20 kHz low frequency component.

Amundin (1991a) used a system whereby he substituted an 80% helium 20% oxygen (heliox) mixture in place of the normal air that a harbour porpoise inhaled. He found that the high frequency component of clicks produced after inhaling the heliox was not shifted in frequency, indicating that this was not produced by gas resonance in the nasal passages. However, the low frequency component was altered by the

introduction of heliox, but the results were complex, with some subjects exhibiting a reduction in peak frequency rather than the expected increase. Amundin suggested that this change indicated the low frequency component to be a resonance phenomenon formed in an air filled cavity, and that the variable frequency change was due to the animal being excited after having the heliox equipment placed over its blowhole. He concluded that the high frequency component was formed in some fixed tissue structure.

It was suggested by Kamminga (1988) that the porpoise's low frequency click component might be used for long range sonar detection, and the high frequency component for short ranges. However, the low source level of the low frequency component together with *Phocoena*'s relative insensitivity to frequencies in the 1-5 kHz region does not appear to support this.

Pilleri *et al.* (1980) (reported in Hatakeyama and Soeda 1990) postulated a transmitted beam pattern of 9° horizontally and 18° vertically from anatomical data taken from a porpoise's skull. However, Hatakeyama and Soeda noted that these values had not been verified experimentally, and they used a beam pattern of 10° both horizontally and vertically in their calculations of Phocoena's echolocation detection capabilities.

Amundin (1991c) investigated the beam pattern of a 2½ year old female *Phocoena* by use of a reference hydrophone placed 4 cm from the animal's rostrum, and a probe hydrophone which could be moved in the plane similarly 4 cm from the rostrum. He measured changes in the amplitude difference between the hydrophones of up to 30 dB from the same click train, which he suggested showed that the animal had some control over the beam pattern. However, these recordings were made in the near field of the porpoise's echolocation beam pattern, where relatively large changes in local intensities may have a minimal effect on the far field beam width. Repetition of the experiment for the far field would be valuable in order to quantify the amount of control the animal has on its transmitted beampattern.

X-ray computer tomography scans of the head of *Delphinus* were made by Aroyan *et al.* (1992), which were used to construct a two-dimensional model of acoustic transmission. They used this model with a finite element method to calculate sound wave paths through the head from a number of proposed sound emitting bodies. The authors discovered a best fit with those echolocation beam patterns recorded from *Tursiops* when they simulated a sound source in the monkey-lip/dorsal bursae complex. Although the prediction used beam patterns from a different species to that of the head, it appears reasonable to expect the sound source to lie in the same structure for both these anatomically similar species.

Echolocation Signal Characteristics

Diercks *et al.* (1973) noted a number of difficulties in accurately recording and measuring the characteristics of echolocation signals. Specifically, the receiving beam pattern and frequency response of the hydrophones must be known, and must be adequate for the task. Similarly, the frequency response of the recording system must also be assessed. If recordings are made off the main axis of the animal's beam pattern, these will show a lower source level and also introduce some spectral coloration, so the orientation of the subject's head must be known. When interpreting

quantitative echolocation data, especially where the orientation of the animals or the characteristics of the recording system were not specified, all the above factors need to be considered.

Watkins (1974) followed up on the work by Diercks *et al.* (1973), and pointed out the difference in attenuation with frequency in water, which would emphasise low frequency components of sounds recorded from distant animals. He also mentions artificial high frequency components which may be produced with overload saturation of amplifiers.

First recordings of the high frequency component of *Phocoena* echolocation clicks were made by Dubrovskii *et al.* (1971). They noted that the previously reported frequency range of 1-3 kHz (Busnel *et al.* 1963, 1965a, 1965b, Busnel and Dziedzic 1967, Dziedzic 1968) would seem to give insufficient target resolution for the porpoise to find its prey of bottom-dwelling fish. Their own recordings showed significant energy for frequencies in excess of 100 kHz (the 3 dB point of their recording equipment).

Møhl and Andersen (1973) came independently to a similar conclusion. They compared the echo level of *Phocoena*'s 2 kHz click component in research with 0.5 mm metal wires, used by Busnel *et al.* (1965b) and Busnel and Dziedzic (1967), with the auditory sensitivity of the porpoise in the same region, and noted a disparity of 100 dB. When Møhl and Andersen used recording apparatus with higher frequency responses than those used by Busnel *et al.* (1965b and 1967), they recorded a significant high frequency component between 100 kHz and 160 kHz, with Source Level averages of 140 dB re. 1 µPa at 1 metre. It appears likely that this source level was an underestimate, due to samples being taken when the animal was pointing anywhere within 30° of the hydrophone, whereas later experiments suggest a transmitted beam pattern much narrower than this (Pilleri *et al.* 1980, cf. Hatakeyama & Soeda 1990). Møhl and Andersen recalculated the echo returned at 130 kHz from the wire for the maximum observed detection range, and found a close match with Phocoena's auditory threshold. They also commented that the maximum energy of the click was approximately one octave above that indicated for best hearing sensitivity from their results (see Voronov and Stosman 1986).

	Value indicated	Source
Echolocation pulse duration	40-200 µsec	Dubrovskii *et al.* 1970
Peak frequency	20-150 kHz	Møhl and Andersen 1973
Source Level (re. 1 µPa at 1 m)	112 dB	*not specified*

Table 1
Phocoena phocoena *echolocation data from Wood and Evans 1980.*
These data appear to have suffered from limitations in the recordings

Data on the echolocation clicks of a number of cetaceans was provided by Wood and Evans (1980), and those presented for *Phocoena* are shown in Table 1. It appears that the problems in recording echolocation signals mentioned by Diercks *et al.* (1973) and Watkins (1974) have probably contributed to the wide ranges given in the above figures for echolocation pulse duration and peak frequency.

A movement with age in the peak frequency of the high frequency component of the click of *Phocoena* was found by Kamminga and Wiersma (1981), who recorded

energies peaking at 140 kHz for a young *Phocoena*, with an adult female producing a lower peak frequency of 120 kHz. This increase for young animals has also been noted by Goodson and Sturtivant (personal observation) with recent recordings of two young porpoises at Harderwijk, Holland, of 8 months and 20 months. The click energy for these animals peaked in the 140-150 kHz region.

Hatakeyama and Soeda (1990) reported typical click pulse widths of 29-83 μs for *Phocoena*, with an inter-click interval of 10-123 ms. They also recorded between 4 and 23 clicks in an echolocation train. The clicks' waveform was reported as containing approximately 9 sine waves, with maximum amplitude generally on the fourth. Hatakeyama and Soeda reported peak frequencies in the 125-140 kHz range for *Phocoena* in a pool, with Source Levels of 158-162 dB re. 1 μPa at 1 metre. These Source Levels are some 20 dB above those reported by Møhl and Andersen (1973), which Hatakeyama and Soeda suggest is due to the large size of the pool and the competition between porpoises when chasing the live fish given to them during the recordings.

Later experiments by Akamatsu *et al.* (1992) involving harbour porpoises in a net enclosure in a fishing port, produced results showing Source Levels still higher at 149 dB to 177 dB. An increase in Source Level during recordings in open water conditions has also been observed in *Tursiops* by Au *et al.* (1974). However, these values for *Phocoena* are still approximately 50 dB below those reported by Au *et al.* for *Tursiops* (215-230 dB re. 1 μPa 1 metre), which may be due to *Phocoena*'s much smaller size.

Evans and Awbrey (1988) noted that clicks produced by *Delphinus* tended to have multi-modal spectra, with strong peaks in the ultrasonic range. Wood and Evans (1980) showed an echolocation waveform and associated spectrum for *Delphinus*, which revealed significant energy in the 10-40 kHz region, peaking at 25.6 kHz, and lower peaks (approximately half the intensity at 25.6 kHz) at 90 kHz and 110 kHz. However, the click spectrum shown on the same page for *Tursiops* indicated a spectral peak at 52 kHz with little energy over 100 kHz.

Although no reference was given for the waveforms, nor any information given about the conditions under which the recordings were made, it appears that these recordings were probably made with captive animals, with information for *Delphinus* possibly coming from those used in Russian research. Recordings of *Tursiops* in the open ocean by Au *et al.* (1974) indicated use of frequencies peaking between 120 kHz and 130 kHz. It seems probable that similar spectra could be used by *Delphinus* in open ocean conditions. Judging by the auditory range of up to 280 kHz reported by Bel'kovich and Solntseva (1970) for *Delphinus*, it would seem that frequencies higher than 40 kHz should form a significant part of the click spectrum.

The data for *Delphinus*'s echolocation signals given by Wood and Evans (1980) are shown in Table 2. Similarly to the data they provided for *Phocoena*, it appears that the quality of the recordings has affected the results obtained, considering the technology available around 1970 in the Soviet Union. The Source Level of 140 dB from Gurevich seems especially suspect. This would place *Delphinus*'s echolocation clicks as quieter than those of *Phocoena* by more than 30 dB (data from Akamatsu 1992). It seems more probable that either reverberant tank conditions caused the animal to use quieter clicks, or that limitations of the recording equipment lead to this low Source Level.

	Value indicated	Source
Echolocation pulse duration	35-350 μsec	Evans 1973
Peak frequency	20-100 kHz	Titov 1972
Source Level (re. 1 μPa at 1 m)	140 dB	Gurevich 1969

Table 2
Delphinus delphis *echolocation data from Wood and Evans 1980.*
These data appear to have suffered from limitations in the recordings

Detection Capabilities

Phocoena was reported by Wood and Evans (1980) to be able to detect 0.2 mm diameter steel wire and 0.9 mm nylon thread (Busnel and Dziedzic 1967), and to detect a 75 mm diameter cylinder of various materials at 8-11 metres (Zaslavsky *et al.* 1969).

Evans (1973) noted that bubbles were seen to form on wires in echolocation trials with *Inia geoffrensis* (Penner and Murchison 1970) which increased the wires' target strengths. Although no mention of bubbles was made in the above similar echolocation trials with *Phocoena* conducted by Busnel and Dziedzic, Evans suggested that this might have resulted in overestimation of *Phocoena*'s echolocation abilities. This fact was again noted by Moore (1980). However, Evans' suggestion of bubbles on the wires was made at a time when its high frequency echolocation component was only just being reported, and may have been an attempt to explain *Phocoena*'s apparently exceptional detection capabilities.

The target strengths of lead-line and netting were determined by Pence (1986) (reported in Hatakeyama and Soeda 1990) to be -33 dB and -55 dB respectively. However, it should be noted that these values were highly dependent on the inclination angle of the net to the echolocation source/receiver. Hatakeyama and Soeda (1990) calculated the detection threshold of the porpoise for a short pulse (43 μs) at 130 kHz to be 94 dB re. 1 μPa. Using the results from Pence, they calculated the maximum detection ranges for lead-line and netting to be 9 m and 2 m respectively.

However, if the netting was inclined at angles of just 3-4° to the source/receiver, they calculated that the detection range would drop to just 60 cm. They noted that in such cases the porpoise would become entangled unless it were moving slowly and carefully or using its vision to some extent. Goodson and Sturtivant (personal observation) have both noted the difficulty of seeing monofilament netting under water, even in bright conditions, and it would appear that vision also may not be of much use in detecting monofilament nets.

Hatakeyama and Soeda (1990) also indicated the importance that variation in range plays in the perceived target strength of large objects (such as fishing nets) when the coverage of the ensonifying beam does not extend to the entire object. They used the equation

$$TS = TS_1 + 10\log(S_r / S_1)$$

to calculate the illuminated target strength in these cases, where TS_1 is the target strength at 1 m, S_r is the area illuminated at a distance r metres from the target, and S_1 is the area illuminated at a distance of 1 m.

Evans and Awbrey (1988) noted that studies with blindfolded *Delphinus* in the Black Sea by Bel'kovich *et al* (1969) showed an echolocation performance at least as good as *Tursiops* at discriminating between complex planar targets. Nachtigall (1980) referred to the work of Bagdonas *et al.* (1970) with *Delphinus*, where a common dolphin was eventually able to discriminate between a 100 cm² square and a 50 cm² triangle using echolocation alone. Work by Bel'kovich *et al.* (1969), again reported by Nachtigall (1980), revealed that *Delphinus* was able to discriminate between the surface area of flat targets in the region of 100 cm² to an accuracy of 10%.

Nachtigall also referred to Bel'kovich and Borisov (1971) who used a discriminatory test with *Delphinus* which required discrimination between a square and a similar square with a hole cut in the middle. The animal attained an 89.5% success rate. Further trials indicated that the dolphin could distinguish between squares with a difference in area of 6.5%. Nachtigall noted that when compared to the earlier trial by Bel'kovich *et al.* (1969), this result would seem to indicate that the subject was using spectral cues as an aid to discrimination in this case.

Problems of Net Detection

Evans *et al.* (1988) (reported by Evans and Awbrey 1988) suggested that the waveform of Dall's Porpoise (*Phocoenoides dalli*), which has a similar structure to that of *Phocoena*, would seem to have adequate resolution for the detection of fishing nets. Unfortunately, the assumptions under which the calculations were made are not stated in Evans and Awbrey (1988). One might easily expect *Phocoenoides* to be able to detect nets at optimal angles, e.g. at precisely 90°, where the summed responses from the length of the filaments would be greatest. However, at angles of just 3°-4°, the target strength of the net can decrease by 10 dB (Pence 1986, referenced in Hatakeyama and Soeda 1990), which might drastically reduce the possible detection range.

Evans and Awbrey (1988) suggested that the high number of net caught animals of *P. dalli* could be explained by the fact that it is usually silent when travelling, and may encounter nets with their sonar off. This assumes that the animals are able to detect the net when using their sonar, over which there is some doubt. However, Goodson and Sturtivant (personal observation) have observed this behaviour of travelling without echolocating in wild *Tursiops truncatus*, both for entire groups of animals, and for stragglers to groups. In this second case, the straggler may use the echolocation clicks from the leaders to indicate in which direction the group is heading. Clearly, echolocating animals will have a greater chance of detecting a net in their path.

Summary

The harbour porpoise and the common dolphin differ substantially in their foraging environments. *Phocoena* appears as an opportunistic feeder, spending much of its time foraging near the sea bed generally on small schooling fish such as herring and

anchovy. *Delphinus* feeds away from the sea floor in less reverberant conditions on pelagic prey such as sardines, anchovies, herring and pilchards.

It appears that *Phocoena*'s limited size influences its maximum echolocation Source Level, and employs a narrow frequency band echolocation pulse. Use of a small frequency range improves the signal to noise ratio of the return, thus increasing the maximum detection range of a low energy pulse, but at the expense of classification potential. Its narrow-band echolocation system appears to be optimised for detection of small prey items at the expense of early target classification.

Delphinus appears, again, to choose a broad range of prey, but employs an echolocation signal with a much wider spectrum than *Phocoena*. This may enable it to use the extra information in the "coloration" of the target echo, due to differences in target size or swim-bladder shape, in order to discriminate prey and non-prey objects soon after detection. Its less reverberant foraging environment, and potentially higher Source Level sonar (although this has not yet been experimentally verified), allow *Delphinus* to use a wide-band echolocation signal.

Delphinus's discrimination capabilities have been compared favourably with those of *Tursiops*, which uses peak frequencies higher than those reported for *Delphinus*. It should be expected that future recordings of this species will show an echolocation signal with significantly increased energy at the higher frequencies and with higher Source Levels than those so far made. *Delphinus*'s sonar appears optimised for early classification of prey in regions near the surface of the ocean.

Differing reports on the auditory sensitivity of *Phocoena* exist. The brain-stem audiogram conducted by Voronov and Stosman (1986) may be more applicable to echolocation ability as this appears to show the response of an echolocation specific receptor. It appears that only a behavioural audiogram exists for *Delphinus*, but it may be that a receptor specific to its peak echolocation frequencies may exist. The behavioural audiogram for *Delphinus* may be more important than for *Phocoena*, as this species appears to make use of its wider click spectrum for object classification by the additional spectral information contained in the returning echo.

The high frequency component of the click for *Phocoena* appears to be produced in some tissue structure in the head, whereas the low frequency component appears to be a gas resonance phenomenon. It is likely that the high frequency click alone is used by the porpoise for echolocation. There is also evidence to suggest that the porpoise may be able to consciously alter the shape of its transmitting beam pattern, although further research is needed in this area.

Little information could be found on the click production mechanism or transmitting beamwidth for *Delphinus*. The best fit for a location of the sound source would appear to be in the monkey-lip/dorsal bursae complex.

The most up to date quantitative information that has been found on the echolocation signals of the species *Phocoena phocoena* and *Delphinus delphis* is summarised in Table 3. The values given for *Delphinus* (marked with an asterisk * in the table) should be treated with caution. The wide range in the peak frequency and click duration, and the low value for the source level, suggest either deficiencies in the recording apparatus used, or recordings made outside the main beam of the animal. More studies need to be published about the echolocation abilities of this species.

When considering detection of large objects, such as fishing nets, the area of coverage of the sonar beam will alter the target strength for returning echoes. Calculations accounting for this suggest that in optimal conditions, *Phocoena* should be able to detect monofilament gill netting at 2 m range. However, the target strength of the net will vary considerably with orientation, and such nets will rarely form a flat surface. In addition, the animal may not encounter the net from the optimal, perpendicular direction. Calculation of detection distances for slightly less than optimal configurations (3-4° away from the perpendicular direction) show a detection range markedly reduced to just 60 cm.

	Phocoena phocoena	*Delphinus delphis*
Peak frequencies	2 kHz and 125-140 kHz (Hatakeyama and Soeda 1990)	20-100 kHz* (Titov 1972)
Click duration	29-83 µsec (Hatakeyama and Soeda 1990)	35-350 µsec* (Evans 1973)
Source level (re. 1 µPa at 1 metre)	149-177 dB (Akamatsu *et al.* 1992)	140 dB* (Gurevich 1969)

Table 3
Summary of current data on echolocation characteristics for Delphinus and Phocoena.
Data marked * may be unreliable due to recording limitations.

Insufficient quantitative data has been gathered about the detection abilities of *Delphinus* to make prediction about detection ranges for this species. However, qualitative data suggests it has similar discriminatory capabilities to *Tursiops*.

Even with enhancements to the target strength of the netting, either by changing the material or by placing discrete reflectors on its surface, if an animal is not using its echolocation when approaching the net, it may not detect its presence until it has blundered into it. However, by obtaining a better understanding of the limitations and capabilities of the sonar systems employed by the species most at risk, one can begin to appreciate the problems they may encounter in detection and classification of nets, and hence provide more effective gear modifications to overcome them.

Acknowledgements

References

Akamatsu, T. *et al.* (1992). "Process of harbour porpoises' entanglement in the gill net". *Technical Report of N.R.I.F.E.*, pp. 25–36 (in Japanese).

Amundin, M., Kallin, E., and Kallin, S. (1988). "The study of the sound production apparatus in the harbour porpoise, *Phocoena phocoena*, and the Jacobita, *Cephalorhynchus commersoni*, by means of serial cryo-microtome sectioning and 3-D computer graphics". In *Animal Sonar, Processes and Performance*, pp. 61–66.

Amundin, M. (1991a). "Helium effects on the click frequency spectrum of the harbour porpoise *Phocoena phocoena.*" *Journal of the Acoustical Society of America* 90(1): 53–59.

Amundin, M. (1991b). "Sound production in odontocetes with emphasis on the harbour porpoise *Phocoena phocoena.*" Doctoral dissertation, University of Stockholm.

Amundin, M. (1991c). "Variable transmission beam pattern in the high frequency click of the harbour porpoise, *Phocoena phocoena.*" Doctoral dissertation, University of Stockholm.

Andersen, S. (1970). "Auditory sensitivity of the harbour porpoise *Phocoena phocoena.*" In *Investigations on Cetacea Vol. II*, pp. 255–259.

Aroyan, J.L., Cranford, T., Kent, J., and Norris, K.S. (1992). "Computer modelling of acoustic beam formation in *Delphinus delphis.*" *Journal of the Acoustical Society of America* 92(5).

Au, W.W.L, Floyd, R.W., Penner, R.H., and Murchison, A.E. (1974). "Measurement of echolocation signals in the Atlantic bottlenose dolphin, *Tursiops truncatus* Montagu, in open waters." *Journal of the Acoustical Society of America* 56(4): 1280–1290.

Au, W.W.L., and Moore, P.W.B. (1988). "The perception of complex echoes by an echolocating dolphin." In *Animal Sonar, Processes and Performance*, pp. 295–299.

Bagdonas, A., Bel'kovich, V.M., and Krushinskaya, N.L. (1970). "Interaction of analysers in dolphins during discrimination of geometrical figures under water." *J. Higher Neural Act.* 20: 1070.

Bel'kovich, V.M., Borisov, I.V., Gurevich, V.S., and Krushinskaya, N.L. (1969). "Echolocating capabilities of the common dolphin (*Delphinus delphis*)." *Zoologicheskii Zhurnal* 48: 876 (in Russian).

Bel'kovich, V.M., and Solntseva, G.N. (1970). "Anatomy and function of the ear in the dolphin." *Zoologicheskii Zhurnal* 49(2): 275.

Busnel, R.G., Dziedzic, A. (1967). "Résultats mestrologiques expérimentaux de l'echolocation chez le *Phocaena phocaena*, et leur comparaison avec ceux de certaines chauve-souris." In *Animal Sonar Systems, Biology and Bionics*, R.G. Busnel (Ed.), Laboratoire de Physiologie Acoustique, Jouy-en-Josas, France, pp. 307–336.

Busnel, R.G., Dziedzic, A., and Andersen, S. (1963). "Sur certaines characteristiques de signaux acoustiques du marsouin *Phocaena phocaena.*" *Compt. Rend.* 157: 245–48.

Busnel, R.G., Dziedzic, A., and Andersen, S. (1965a). "Seuils de perception du système sonar du marsouin *Phocaena phocaena* L." *C.R.Acad.Sc.Paris* 257: 245–48.

Busnel, R.G., Dziedzic, A., and Andersen, S. (1965b). "Rôle de l'impédance d'un cible dans le seuil de sa détection par le système sonar de marsouin *Phocaena phocaena.*" *Compt. Rend. Soc. Biol.* 159(1): 69–74.

Diercks, K.J., Trochta, R.T., and Evans, W.E. (1973). "Delphinid sonar: Measurement and analysis." *Journal of the Acoustical Society of America* 54(1): 200–4.

Dziedzic, A. (1968). "Quelques performances des système de détection par echos des chauves-souris et des delphinidae." *Rev. d'Acoust.* 1: 23–28.

Dubrovskii, N.A., Krasnov, P.S., and Titov, A.A. (1970). "On the problem of emitting ultrasonic echolocation signals by harbor porpoise (*Phocoena phocoena*)." *Akusticheskii Zhurnal* 4: 521–25 (in Russian).

Dubrovskii, N.A., Krasnov, P.S. and Titov, A.A. (1971). "On the emission of echolocation signals by the Azov Sea harbour porpoise." *Soviet Physics Acoustics* 16(4): 444–48.

Evans, W.E. (1973). "Echolocation by marine delphinids and one species of freshwater dolphin." *Journal of the Acoustical Society of America* 54(1): 191-99.

Evans, W.E, and Awbrey, F.T. (1988). "Natural history aspects of marine mammal echolocation: Feeding strategies and habitat." In *Animal Sonar, Processes and Performance*, pp. 521–34.

Evans, W.E., Awbrey, F.T., and Hackbarth, H. (1988). "High frequency pulses produced by free ranging Commerson's dolphin (*Cephalorhynchus commersonii*) compared to those of phocoenids". *Report of the International Whaling Commission* (Special Issue 9), pp. 173–81.

Gurevich, V.S. (1969). "Echolocation discrimination of geometric figures in the dolphin, *Delphinus delphis.*" *Vestnik Moskovskogo Universiteta, Biologiya, Pochovedeniye* 3: 109–12 (in Russian).

Hatakeyama, Y. and Soeda, H. (1990). "Studies on echolocation of porpoises taken in salmon gillnet fisheries." In *Sensory Abilities of Cetaceans, Laboratory and Field Evidence,* pp. 269–81.

Hohn, A.A. (1990). "Harbour porpoise." In *Whales and Dolphins,* 170–71.

Jefferson, T.A. and Leatherwood, S. (1990). "Common dolphin." In *Whales and Dolphins,* 146–47.

Kamminga, C. and Wiersma, H (1981). "Investigations on cetacean sonar II, acoustical similarities and differences in odontocete sonar signals." *Aquatic Mammals* 8(2): 41–62.

Kamminga, C. (1988). "Echolocation signal types of odontocetes." In *Animal Sonar, Processes and Performance,* pp. 9–22.

Møhl, B. and Andersen, S.A. (1973). "Echolocation; High-frequency component in the click of the harbour porpoise." *Journal of the Acoustical Society of America* 54: 1368–73.

Moore, P.W.B (1980). "Cetacean obstacle avoidance." In *Animal Sonar Systems,* 97–108.

Nachtigall, P.E. (1980). "Odontocete echolocation performance on object size, shape and material." In *Animal Sonar Systems,* pp. 71–95.

Pence, A.E. (1986). "Monofilament gill net acoustic study." Applied physics lab., University of Washington, 13 pp.

Penner, R.H. and Murchison, A.E. (1970). "Experimentally demonstrated echolocation in the Amazon river porpoise, *Inia geoffrensis* (Blainville)." *Naval Undersea Center Tech. Publ. No.* 187.

Pilleri, G, Zbinden, K., and Kraus, C. (1980). "Characteristics of the sonar system of cetaceans with pterygoschisis, directional properties of the sonar clicks of *Neophocaena phocaenoides* and *Phocoena phocoena* (Phocoenidae)." In *Investigations on Cetacea Vol. XI,* pp. 157–88.

Popper, A.N. (1980). "Behavioural measures of odontocete hearing." In *Animal Sonar Systems,* pp. 469–82.

Poppov, V.V., Ladygina, T.F., and Supin, A.Ya. (1986). "Evoked potentials of the auditory cortex of the porpoise, *Phocoena phocoena.*" *Journal of Comparative Physiology A* 158: 705–11.

Schevill, W.E., Watkins, W.A., and Ray, C. (1969). "Click structure in the porpoise *Phocoena phocoena.*" *Journal of Mammalogy* 50(4): 721–28.

Sukhoruchenko, M.N. (1973). "Frequency discrimination of the dolphin *Phocoena phocoena.*" *Fiziologicheskiy Zhurnal imeni I.M. Sechenova* 59(8): 1205 (in Russian).

Titov, A.A. (1972). "Investigation of the sound activities and peculiarities of the Black Sea dolphin's echolocator." PhD Thesis, Karadag (In Russian).

Turl, C.W. (1993). "Low-frequency sound detection by a bottlenose dolphin." *Journal of the Acoustical Society of America* 94(5): 3006–9

Voronov, V.H. and Stosman, I.T. (1986). "Electrical responses of the stem structures of the acoustic system of *Phocoena phocoena* to tonal stimuli." In *The Electrophysiology of the Sensory Systems of Marine Mammals,* V.E. Sokolov (ed.), Nauk, Moscow (in Russian).

Watkins, W.A. (1974). "Bandwidth limitations and analysis of cetacean sounds, with comments on 'Delphinid sonar: Measurement and analysis' [K.J. Diercks, R.T. Trochta, and W.E. Evans, J. Acoust. Soc. Am. 54, 200-204 (1973)]." *Journal of the Acoustical Society of America* 55(4): 849–53.

Wiersma, H. (1988). "The short-rime-duration narrow-bandwidth character of odontocete echolocation signals." In *Animal Sonar, Processes and Performance,* pp. 129–45.

Wood, F.G. and Evans, W.E. (1980). "Adaptiveness and ecology of echolocation in toothed whales." In *Animal Sonar Systems,* pp. 381–425.

Zaslavsky, G.L., Titov, A.A., and Lekomtsev, V.M. (1969). "Investigations of the echolocation abilities of the harbor porpoise (*Phocoena phocoena*)." *Trudy Akusticheskogo Instituta* **8**: 134–8.

Bibliography

Animal Sonar, Processes and Performance (1988). Paul E. Nachtigall and Patrick W.B. Moore (Eds.); Plenum Press, New York and London; ISBN 0-306-43031-2.

Animal Sonar Systems (1980). René-Guy Busnel and James F. Fish (Eds.); Plenum Press, New York and London; ISBN 0-306-40327-7.

Echolocation in whales and dolphins (1983). Purves, P.E. and Pilleri, G.E., Academic Press, New York.

Sensory Abilities of Cetaceans, Laboratory and Field Evidence (1990). Jeanette A. Thomas and Ronald A. Kastelein (Eds.); Plenum Press, New York and London; ISBN 0-306-43695-7.

Whales and Dolphins (1990). A.R. Martin (ed.); Salamander Books Ltd., London and New York; ISBN 0-86101-488-X.

Appendix B : Implementation Details

B.1 The Fourier Transform

B.1.1 Fast Fourier Transform Background Theory

For a continuous function $h(t)$ in the time domain, the Fourier transform $H(f)$ of the function gives the representation of the function in the frequency domain. $H(f)$ can be calculated from $h(t)$ using the following formula:

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{2\pi ft} dt$$

For a discrete function $h(t_k)$ with N values sampled consecutively with a uniform sampling interval Δ, so that $t_k = k\Delta$, for $k = 0, 1, 2, \dots N-1$,

$$H(f_n) = \int_{-\infty}^{\infty} h(t) e^{2\pi f_n t_k} dt \approx \sum_{k}^{N-1} h_k e^{2\pi f_n t_k} \Delta = \Delta \sum_{k=0}^{N-1} h_k e^{2\pi i kn/N}$$

where $h_k \equiv h(t_k)$. The discrete Fourier transform is defined as the summation term shown at the end of the equation. If we denote this as H_n, then

$$H_n \equiv \sum_{k=0}^{N-1} h_k e^{2\pi i kn/N}$$

and the formula for the discrete inverse Fourier transform is

$$h_k = \frac{1}{N} \sum_{k=0}^{N-1} H_n e^{-2\pi i kn/N}$$

The *fast* Fourier transform technique takes advantage of the following property:

$$H_k = \sum_{j=0}^{N-1} e^{2\pi i jk/N} h_j$$

$$= \sum_{j=0}^{N/2-1} e^{2\pi i jk/(N/2)} h_{2j} + W^k \sum_{j=0}^{N/2-1} e^{2\pi i jk/(N/2)} h_{2j+1}$$

$$= H_k^e + W^k H_k^o$$

where $W = e^{2\pi i/N}$, H_k^e denotes the kth component of the Fourier transform of length $N/2$ formed from the even components of the original f_j's, and H_k^o similarly for the odd components (Danielson-Lanczos Lemma). This property can be used recursively to calculate H_k^e and H_k^o:

$$H_k^e = H_k^{ee} + W^k H_k^{eo}$$
$$\text{and } H_k^o = H_k^{oe} + W^k H_k^{oo}$$

which eventually reaches the termination condition $H_k^{eoeeoeo...oee} = h_n$ for some n. This recursive system can be implemented in a simple way on a list of h_k's by reordering the elements in the list. The reordering is accomplished by taking the index of the element in the list (in the range 0 to $N-1$) and reversing the bits in the index. This manipulation groups all of the odd elements in the first half of the list and all the even in the second. Furthermore, for each half of the list all the odd elements *in that half* appear in the first half and the even elements in the second half. Thus, for the condition where $H_k^{eoeeoeo...oee} = h_n$,

$$H_k^{eoeeoeo...oe} = H_k^{eoeeoeo...oee} + W^k H_k^{eoeeoeo...oeo}$$

$$H_0^{eoeeoeo...oe} = h_{eoeeoeo...oee} + h_{eoeeoeo...oeo}$$

$$H_1^{eoeeoeo\ oe} = h_{eoeeoeo\ .oee} - h_{eoeeoeo...oeo}$$

and $$H_k^{eoeeoeo...oo} = H_k^{eoeeoeo...ooe} + W^k H_k^{eoeeoeo...ooo}$$

$$H_0^{eoeeoeo...oo} = h_{eoeeoeo..ooe} + h_{eoeeoeo...ooo}$$

$$H_1^{eoeeoeo\ oe} = h_{eoeeoeo...ooe} - h_{eoeeoeo...ooo}$$

In both above cases, the pair of h_k elements in each equation are adjacent after reordering. Thus, the algorithm can move through the reordered list operating on pairs of data. For the next case up, where $N=4$, the following equations hold:

$$H_k^{eoeeoeo\ o} = H_k^{eoeeoeo..oe} + W^k H_k^{eoeeoeo...oo}$$

$$H_0^{eoeeoeo...o} = h'_{eoeeoeo...oee} + h'_{eoeeoeo...ooe}$$

$$H_1^{eoeeoeo.\ o} = h'_{eoeeoeo...oeo} + h'_{eoeeoeo...ooo}$$

$$H_2^{eoeeoeo.\ o} = h'_{eoeeoeo...oee} - h'_{eoeeoeo...ooe}$$

$$H_3^{eoeeoeo.\ o} = h'_{eoeeoeo...oeo} - h'_{eoeeoeo..ooo}$$

where h' indicates the result from the previous iteration ($N=2$). A symmetry of the process has been used in the above equations and in the algorithm itself: $W^{N/2} = -1$, therefore $W^{k+(N/2)} = -W^k$. Since all pairs processed are separated by $\frac{\pi}{2} = W^{N/2}$, this symmetry can reduce calculations significantly.

The iteration is carried out until the values for the complete transform is achieved (i.e. $H_k = H_k^e + W^k H_k^o$)

B.1.2 Outline of the FFT Algorithm

The FFT algorithm takes as arguments a 1-dimensional array of complex values and the number of elements in the array. The complex values were represented for convenience as a structure containing floating point values for the real and imaginary components.

The first stage in the algorithm was to reorder the data (termed 'decimation') by reversal of the indices of the elements. This involved reversing the bits of the

index of the element to be reordered in the array, which was achieved as follows (in routine rev_bits() in the file fft.c):

Set the bit reversed result to 0
Do loop number-of-bits times
Bit shift the result by one place to the left (double value)
If lowest bit of number-to-reverse is 1, then set lowest bit of result to 1
Bit shift the number-to-reverse by one place to the right (halve value)
End loop
Return the result

This algorithm takes advantage of the bit shifting ability of the processor, and avoids the use of any, more costly, arithmetic operations. Clearly this process is symmetrical, with the reversal of a reversal returning the original number. Thus the indexed element and its bit reversed counterpart may simply be exchanged. In its pass through the data, the reordering algorithm needed to exchange only those element pairs which had not already been swapped. This was resolved by only exchanging those elements where the bit reversed index was greater than the original. Of course, when the bit reversed index was the same as the index itself, it was not necessary to carry out a swap as the element was already in its correct place.

The number of bits to be reversed was not passed to the FFT routine, and so was required to be calculated. A description of the calculation of the number of bits in the transform is given below (used in routines fft() and ifft() in the file fft.c):

Set check-value to 2
Loop with number-of-bits initially set to 1, incrementing number-of-bits each loop
Bit shift check-value to the left by one bit (doubles the value)
Loop if check-value still less than number-of-elements

For the FFT algorithm to operate correctly, the number of elements must be a power of two. If this condition does not hold, the data ought to be padded out in some way to the next appropriate power. For the routine which converted from the raw signal to the FFT representation, if there was insufficient data to complete a full partition for the FFT (due to an end-of-file occurrence) then the partition was discarded.

The main body of the FFT algorithm is presented below (implemented in routine fft() in the file fft.c). Note that array indices are taken as starting from 0 in 'C', often simplifying array access calculations.

Loop with Element Pair Step starting from 1 but always less than Array Size (Element Pair Steps is half the partition size for calculation purposes)
Set Theta to p / Element Pair Steps (equivalent to 2p/N)
Set the real part of W Increment to cosine of theta
Set imaginary part of W Increment to sine of theta

Set W to 1.0

Loop through First Element Offset from 0 but less than Element
Pair Steps (all pairs in a partition)
Loop through First Element index starting at First
Element Offset and increasing by Element
Pair Steps * 2 (moves through all partitions)
Set Second Element index to First Element index
+ Element Pair Step

Set Product to complex product of Second
Element and W
Add Product to First Element
Subtract Product from Second Element
Loop on First Element index until it exceeds Array Size
Multiply W by W Increment (complex form)
Loop on First Element Offset until it reaches Element Pair Steps
Loop on Element Pair Step, multiplying it by 2 until it reaches Array
Size

This algorithm utilises the Danielson-Lanczos Lemma for the Fourier
transform explained in the previous section. After decimation of the original data list
using the bit reversal routine, the full Fourier transform can be constructed by firstly
operating on all pairs of elements (Element Pair Step = 1), then operating on pairs
of pairs, etc., until the partition includes the entire data array.

B.2 Choice of the Colour Palette

A system was required which would display the required range of values
giving the maximum differentiation possible between close values without causing
confusion as to the relative power levels. The logarithm to the base 10 of the power
was used to identify which colour should be drawn on the screen, which enabled
visual identification of quiet whistles in the presence of much louder echolocation
clicks.

A system of representing increasing power by increasing colour intensity was
first tried using the existing palette colours, where the hue of the colour was not
significant. Although this allowed good differentiation between close values, the
arbitrary hue choice gave a confusing impression and did not aid in identification of
absolute power values.

A second system was examined which consisted of a grey scale. This proved
to have the opposite effect, producing good indications of the power value, but
differentiation between adjacent power levels proved difficult.

The third system was the one which was believed to give the best results, and
has been used throughout all the utilities when an FFT display was required. There
are three commonly used models for defining colours in computer graphics — RGB
(red, green, blue), HSV (hue, saturation, value), and HLS (hue, lightness, saturation).
Computer painting packages often use either HSV or HLS models for allowing the
user to define colours, as they allow a more intuitive method of specifying colours
than the RGB system. The HLS system was used in the definition of the FFT colour
palette, and a short description of its use is needed.

In the HLS model, the 'hue' H specifies what one may call the 'basic colour', i.e. red, green, violet, magenta, orange, etc.. It is specified as an angle on a 'colour circle' with red, green, and blue each separated by 120°. Figure 1 indicates the relative intensities at each angle:
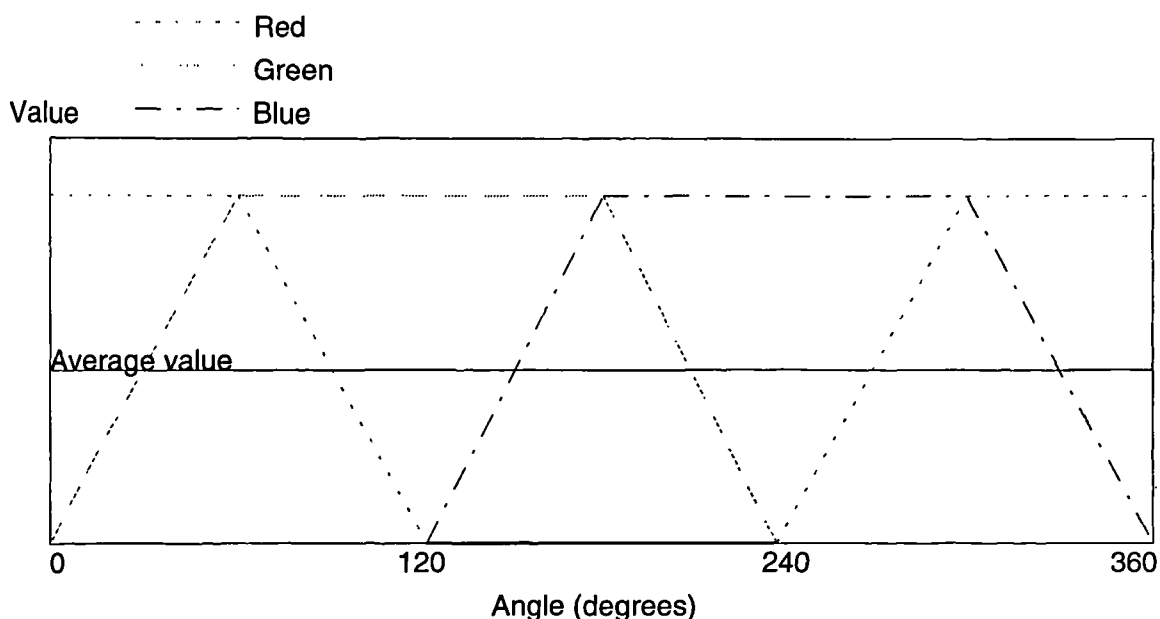

Figure 1 - Relationship of component colours for the Hue parameter

As can be seen, values at 0° give pure red, at 60° yellow, at 120° green, at 180° cyan, at 240° blue, at 300° magenta, and then returning to red at 360°. The colours are interpolated between these six values.

The 'lightness' L gives the brightness of the shade, ranging from 0.0 (black) to 1.0 (white). Maximum lightness is generally equated to the brightest white displayable, e.g. for a 24 bit display, a lightness of 1.0 would be red, green, and blue values all set to 255 in the common RGB specification. Lightness is often represented vertically, with black at the bottom, and white at the top.

The 'saturation' S of the colour is the contribution made by the hue away from neutral grey. Thus a saturation value of 0.0 would produce a grey with a brightness depending solely on the lightness L, whereas low values of saturation would give 'pastel' shades, and high values very bold colours.

Although hue and lightness can vary in value between 0°-360° and 0.0-1.0 respectively, in practice the saturation is limited by the value of the lightness. For instance, at L=0.0 (black) saturation S must also be zero, as any value of S would require at least one colour to have a 'negative' intensity value. Similarly at L=1.0 (white), saturation again must be zero as all RGB colour components are already at maximum. This leads to the view of the HLS model as a double cone shown in Figure 2, with lightness on the vertical axis, saturation perpendicular to this axis, and hue specifying the direction of the saturation vector.
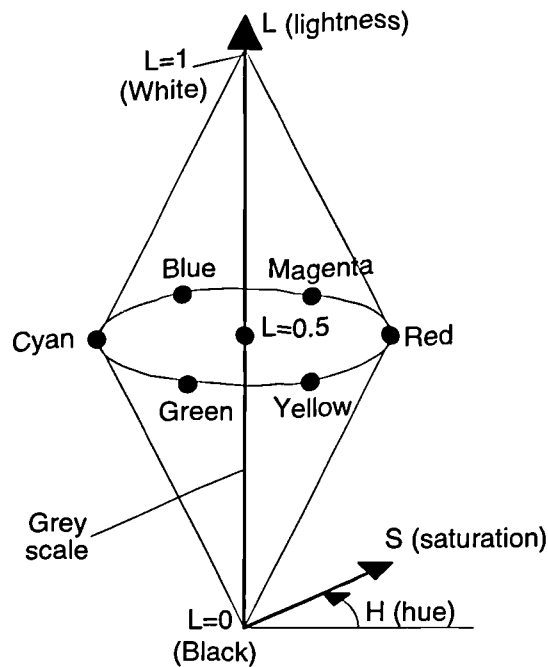
Figure 2 - Schematic of the HLS colour model

In order to produce maximum differentiation between adjacent values and also to maintain a smooth and intuitive display of the magnitude of the power, the colour palette was initialised using luminance values between 0.0 and 1.0, hues starting at 0° and moving through the full circle, and saturation values at maximum. This produced a palette which used dark blue and green colours for low intensity values, through towards brighter yellows and reds for higher values, and bright pinks and white for the highest values.

B.3 The Three-dimensional Viewing Programme

B.3.1 Calculation of the Surface Normal

After transforming the points according to the oblique projection parameters, the surface normal to each triangular facet was calculated using the cross product of vectors between two vertices taken in a clockwise manner. For triangle vertex position vectors $\overline{A}$, $\overline{B}$, and $\overline{C}$, forming a clockwise group when viewed from above, the surface normal $\overline{n}$ can be calculated by

$$\overline{n} = \overline{p} \wedge \overline{q} \quad where \quad \overline{p} = \overline{B} - \overline{A}, \quad and \quad \overline{q} = \overline{C} - \overline{B}$$

Evaluation of the y-component of the triangle's surface normal $\overline{n}$ indicates whether the triangle is visible to the viewer. If the y-component is positive, then the surface faces away from the viewer, and is therefore not visible and does not need to be drawn. This allows significant drawing time to be saved by making this calculation and ignoring all triangles which are not visible.

B.3.2 The Phong Lighting Model

The Phong lighting model (Phong 1975), a widely used model in computer graphics, was used to calculate the lighting effects. This model calculates the

B-6

intensity at a pixel for each red, green, and blue colour component based on the ambient lighting, diffuse reflection, and specular reflection at a point on a surface. The different components arise as follows:

Ambient lighting is assumed to be uniform from all directions, and is usually included to avoid a 'harsh' appearance to the rendered scene.

The diffuse reflection component arises from the illumination due to some point light source, and will be dependent on the intensity of the light source, the distance from source to surface, the angle between the surface and the illuminating rays, and the illumination pattern of the source (although this is often assumed to be uniform for ease of calculation). The diffuse component is scattered in all directions uniformly.

The specular reflection is the component arising from the mirror-like properties of the surface, and involves calculating the 'ideal' reflected ray and a distribution around it to enable variation in the 'mattness' or 'glossiness' of the surface to be shown. The specular reflection takes the colour of the light source alone, whereas the diffuse reflection is coloured by both the light source and the surface colour.

The lighting model used in the 3d() programme used only the ambient and diffuse properties of the full Phong model, as the specular component requires a large amount of calculation compared to the other two components. The effect of this on the rendered image is to give all surfaces a matt finish, which in this case does not detract from the overall impression.

One light source was used, which was assumed to be uniform, white, and at infinity, thus simplifying the equations. The colour of the surface could then be found as the product of some scaling factor and the base colour of the surface, as no colour component additions needed to be made due to specular reflection taking the light source colour. The resultant intensities I_λ for $\lambda=\{$red, green, blue$\}$, could be calculated from

$$I_\lambda = C_\lambda (I_a + I_p(\overline{n} \cdot \overline{l}))$$

where C_λ are the red, green, and blue coefficients of the surface colour, I_a is the ambient lighting intensity, I_p is the intensity of the point light source at the surface (assumed to be a constant in this case), $\overline{n}$ is a unit normal to the surface, and $\overline{l}$ is a unit vector pointing toward the light source. It is normal for the sum of the ambient, diffuse, and specular lighting intensities to be 1.0, and if this is the case the colour components (C_r, C_g, C_b) can be taken as the RGB colour values of the surface, giving the resulting RGB colour to use for drawing the surface as (I_r, I_g, I_b).

The palette was initialised with 64 shades of each of the 4 base colours, giving a total colour range of 256. This produced a pleasing effect which was similar to that of a mountain range in an ocean. Whistles were easy to pick out due to their extended time and slowly varying frequency, causing them to be shown as horizontal ridges across the screen. Even when the signal to noise ratio was poor, it was often possible to visually identify faint whistles due to the long 'escarpment' that they produced.

B.4 The Minimum Spanning Tree Algorithm

B.4.1 The Minimum Spanning Tree

The minimum spanning tree (MST) algorithm has a number of applications to problem solving in computer science, although its order N^2 execution time, where N is the number of data nodes, often restricts it to off-line use. The MST is constructed from a non-directed graph, containing a set of nodes, and a list of weighted edges between the nodes. To move from one node to another (when the nodes are connected by a linked set of edges) will 'cost' the sum of the weights of the edges used. Such a graph will normally contain loops, where it is possible to follow a linked list of edges and arrive back at the start node without having used any edge more than once. Indeed, if the graph does not contain any loops then calculation of the MST is pointless, as there would be no alternative routes to be considered.

The MST provides a tree which contains a subset of the edges from the original graph. It has the property that if a path is traced through the MST between any two points, including initial travel up the tree, then that path will have a lower 'cost' than for any other route not contained in the MST. There is a variation of this definition which is particularly applicable to pattern recognition and cluster analysis. This is when the distance from a node to a sub-tree is the shortest distance between the node and any node existing in that sub-tree. In this case the sub-tree is acting similarly to a cluster, and the distance calculated is the distance of the test point from the boundary of the cluster. This model also holds for a test point in the middle of a toroidal cluster, or similar volumes which often prove difficult to analyse.

B.4.2 MST Implementation

The cluster modification was used when implementing the MST. Each node in the tree consisted either of a link to two other nodes, or of a leaf indicating a point in the data set. For a link node, the two sub-nodes were considered as sub-clusters, and a length was given for the link node which was the minimum distance between the two sub-clusters. A further condition was that the length for the link node was greater than all lengths in either of the two sub-clusters. Thus the MST produced contained N leaves (indicating data points on the FFT) and N-1 link nodes (indicating distances between subclusters), giving a total of $2N$-1 data elements.

The length between two data points was calculated as the sum of the squares of the differences in x-, y- and z-values for the two points. An additional weighting was provided for each axis in order to provide a bias for:

patterns extended in the time domain (smaller x-difference weighting), such as whistles,

patterns extended in the frequency domain (smaller y-difference weighting), such as clicks,

patterns with similar amplitudes (larger z-difference weighting).

The basic algorithm is as follows:

The two points with the smallest distance between them must be included in the MST, and so these two points are included in the tree as leaves, separated by a link node

which contains the length between the points, and pointers to the leaves. This link node is termed the root node, and the MST is now a cluster of two points.

The smallest distance between any point contained in the MST and a point not in the tree is determined.

This new point becomes a sub-cluster, and a link node is placed in the MST so that the new point forms one sub-cluster and the point that it was closest to is contained in the other sub-cluster. The link node is also positioned so that there are no links above it in the MST which contain a shorter length than its own.

If the new link node has the root node as one of its sub-clusters, then the new link node becomes the root node.

If there are still points not represented in the MST, then the process returns to step 2.

Thus the MST consists of leaves indicating points, forming sub-clusters of one element, and link nodes indicating distances between the two sub-clusters below them and also forming a new subclusters. This MST can then be searched to find those sub-clusters which contain maximum lengths below a certain threshold. Each of these sub-clusters could then be categorised. Small clusters could be labeled as noise, long narrow clusters as tonal signals (possibly whistles), and tall thin ones as impulsive sounds (possibly clicks). There were, however, a number of practical and potential problems with this method.

B.4.3 Optimisations Made to the Basic MST Algorithm

Early versions of the MST algorithm took up large quantities of time searching for points either inside or outside the MST. A number of optimisations were made, most of which had the effect of reducing execution time at the expense of an increase in the memory requirements of the programme. This section contains a list of the optimisations which were implemented.

The search for the initial pair of points to start the MST was very time consuming as all combinations ($N^2 - N$ pairs) were compared. Often, when areas of the data array fell below the minimum cut-off value, these areas were given the same zero z-coordinate. In this case, the smallest possible length would be encountered, being the sum of the x- and y-difference weightings. A condition was added that if this minimum value was encountered during the initial search then the search could be terminated. This greatly decreased the time taken to find an initial 'seed' pair for the MST.

As it stood, the basic algorithm required the testing of lengths between all possible pairs with one point in the MST and one outside the MST. It was considered that this search area could be reduced by defining a search region around the first point in the pair. When the first point was selected, only those other data points within a maximum distance in their x- and y-coordinates would be searched to determine a minimum length. This maximum search range was made accessible to the user. Clearly a maximum distance of 1 would required up to 8 length calculations, rather than for each of the typically 128x127 points in the entire table, and this would significantly increase execution speed. However, whistles may be interrupted by clicks which might cross their signal for one or two time periods, so setting this maximum distance to so low a value could adversely affect the results. Setting the

maximum distance to a larger value, say 3 (giving 48 calculations per point in the MST), would reduce these adverse effects.

Although not implemented, a more general rule was found which would allow a smaller area to be searched without the possibility of missing the smallest distance. For each node in the MST the boundary of squares of incrementing width would be searched. Regarding only the x- and y-coordinate differences, if the sum of the squares of the distances (modified by the weightings for those axes) was greater than the current smallest distance found, then that point could be skipped as it was impossible for it to be closer when its z-coordinate was included in the distance calculation. If all the data points in the square failed this x-y distance test then the next point in the MST could be tested, as there would be no possibility that any points further away could have a smaller distance measure.

Considerable time could be spent in searching the MST to determine whether a point was already in the tree. This became a major factor in execution speed as the MST grew in size during calculation. A flag array was used to indicate whether the point was inside or outside the MST. Initially all points were indicated as outside the MST, then as points were entered they were flagged as contained by the MST in the flag array.

In the later execution of the algorithm when the MST had increased to contain a considerable number of points, it became apparent that each pass the algorithm was checking points contained in the MST which were surrounded by points also contained in the MST up to the maximum searchable distance set by the user. During the searching process, if a MST point was encountered which had no point outside the MST to which it could be connected, then the point was marked as 'dead' and was ignored by further searches. This optimisation had a noticeable effect when the MST had grown to include approximately 25% of the available points, increasing execution speed by around four times.

B.4.4 Comments on the MST

As mentioned previously, the main drawback found with the MST was the time taken in its calculation. In this implementation memory usage was also a problem, although this could be remedied by using a file on disk (possibly on a RAM disk) to hold the intermediate tree in some format, but again at the expense of execution time. Subsequently a 32-bit 'C' compiler has been acquired which allows access to the full 8 Mb available on the system, although many of the problems with memory capacity still hold for some of the larger test samples.

During construction and optimisation, as the operation of the MST became more fully appreciated, it seemed that the cluster based MST approach may not be as useful as was at first thought. A scenario arose where a whistle may emerge gradually from the background noise. If the minimum threshold is used, whereby all (offset) values below a zero were set to zero (as used in the 3d() programme), then the sub-cluster containing these 'floor' values will be very close to those constituting the beginning of the whistle, which in turn will be close to those further along the whistle, etc.. In this case, a sub-cluster which was classified as a 'floor' noise level cluster may also contain the whistles being searched for.

A way around this would be to remove the 'floor' value from the calculations. This would have the effect that determination of the initial 'seed' pair of points for the MST would in all probability require a search through the entire data set, which would be costly and again give an increase in execution time. This approach works because the z-axis values are calculated from the frequency intensity in dB, thus low intensity noise would produce a rough surface, and higher intensity whistles would not have the same roughness due to noise.

It is believed that the minimum spanning tree still represents a powerful way to extract information from the FFT data. However, due to the large computation time for isolation of a single whistle, faster methods for feature extraction were developed, although it is possible that these may not be so precise as use of the MST. If more computational power becomes available at a later date, and restrictions on the time duration become less important, then the MST algorithm will be developed further.

B.5 The First Tracing Algorithm

The original algorithm was designed to search recursively all possible routes across the FFT display and select the one which produced the highest sum of intensities (in dB). However, the time taken to compute this route was excessive. A typical FFT array chosen as a good test case contained 128 frequency bins, and 127 time partitions. For a maximum jump of 1 frequency bin between partitions (which was by no means guaranteed to be able to follow the whistle) the algorithm would be required to search through approximately $128 \times 3^{126} = 1.677 \times 10^{62}$ (ignoring reductions due to traces encountering the top or bottom of the array) which is too large a number of calculations to be comfortably imagined!

Two modifications were made to the routine that reduced the number of calculations required, at the expense of generality. An offset was applied to each intensity value to reduce its value, which had the effect of defining the average intensity of the trace required for a valid whistle. A second value was also introduced which was a threshold value. If the running sum fell below this threshold value, then the trace was stopped. By manipulating these two parameters, the amount of low intensity values which would be allowed before rejecting the trace, and the expected average level of the whistle, could both be altered to find the optimum trace.

The trace with the highest summation when it reached the right hand side, or failing this the trace which extended furthest towards the right hand side, would be chosen as the one which contained the whistle. The beginning and end points of the whistle could be calculated by examination of the intensity values across the trace, with end points defined at some minimum signal to noise ratio.

In order to reduce the amount of calculation time, when the summation from a point to the right hand edge of the screen had been calculated (or as far to the right as possible), the best summation was stored for that point. Subsequently, should any other trace encounter the same point, it could be determined whether the summation would fall below the threshold value, and if so the point could be ignored.

B.6 The Second Tracing Algorithm

This algorithm could be considered as containing two distinct parts: detection of the start of the whistle; and whistle tracing. The initial detection of the whistle was done by comparing the value for each frequency bin with a minimum value specified by the user. If no frequency bin in a partition exceeded this value, then all values in the partition were set to zero, thus indicating that no whistle was present. When a maximum intensity in a partition was found which did exceed this minimum value, that 'seed' point f_s was used as a starting point for the whistle tracing part of the algorithm. Clearly, if the user were to choose too large a value for determining the 'seed' point, then part or all of the whistle could be lost.

The user specifies the allowable change in frequency bins Δf from one time partition to the next. The range of frequency bins from $f_s + \Delta f$ to $f_s - \Delta f$ was searched to find the bin f_m with the highest value. If the ratio of the maximum value found in this range to the minimum value found was less than a user-specified minimum allowable local 'signal-to-noise' ratio, then the trace was terminated at that point. This corresponds to a whistle either finishing or being indistinguishable from the background. Potentially, this may cause problems with whistles which momentarily are indistinguishable from the background, but this did not occur with the test data used.

If the whistle was not indicated as terminating, then all the frequency bins outside the range $f_m + \Delta f$ to $f_m - \Delta f$ were set to zero, and the value f_m was used as the seed frequency bin f_s for the subsequent partition.

This procedure produced a modified FFT which had the identified maximum whistle frequency bin centred in a range of $2\Delta f + 1$ non-zero bins, with all other frequencies set to zero. It should be noted that the central whistle frequency bin cannot be assumed to be the maximum value of a time partition, as in general $f_m \neq f_s$ so not all of the bins in the range $f_m \pm \Delta f$ will have been searched.

Appendix C: Dolphin Whistle Classification with the 'Dolphin' Software

C. Sturtivant & S. Datta

1. INTRODUCTION

Identification and re-identification is often an important part of field studies of dolphins. This task has traditionally been accomplished with photographic identification techniques, although it is also possible to use acoustic methods to obtain the same information. The hypothesis that dolphins have a 'signature' whistle was proposed over thirty years ago [1], and suggested that the majority of dolphin whistles carried the identity of the vocalising animal. Although recently questioned [2], the more current proposal still maintains that whistles carry identity information, but in the wider context of groups of animals. Such identifying whistles have been found for several dolphin species as well as for the killer whale [3,4,5,6,7], and so this method of identification potentially has applications to a wide number of toothed cetaceans.

The Underwater Acoustics Group at Loughborough University has developed software, named 'Dolphin', for quantitative comparison of dolphin whistles [8,9,10]. The software can identify parts of recordings containing whistle-like sounds, extract their frequency-time-intensity contours, and then apply automatic pattern recognition techniques to classify them against previous whistles. These techniques have the benefits of being both objective and quantitative, and provide a probability that any candidate whistle belongs to each existing class, or to some new class.

The mathematical background to the software routines has been presented elsewhere in these proceedings [11]. This paper will describe and explain the use of the 'Dolphin' program, with an example of how a typical whistle contour is extracted from background noise, encoded, and classified.

2. HARDWARE REQUIREMENTS

The software was written to execute on a standard IBM PC, with no additional custom hardware. The platform that is currently used is a 133 MHz PCI Pentium running Windows 95, with a Matrox Mystique graphics card and a Sound Blaster 16 card. The development system for the software was a 486 DX 50 with Sound Blaster 16 card, although the software is designed to run on a platform containing an Intel 386 processor or later.

The development environment was the Gnu GCC C++ compiler, ported to the PC as DJGPP, and the GRX graphics extensions. The compiler produces 32-bit code, and was chosen since this runs considerably faster than 16-bit code required for pre-386 processors. Many of the calculations made in the program are computationally intensive, and it is doubtful that processors slower than a 486 would be suitable hosts for the software.

The choice of the Sound Blaster 16 as the sound card for the software was due to its wide commercial availability, and its ability to sample at sufficiently high sample rates (44 kHz) for dolphin whistles and at a suitable resolution (16 bits). Although better sound cards are readily available today, this seemed the best choice at the beginning of the project. The software is written modularly, so that should support for another sound card be required it could easily be incorporated into the program.

3. SOFTWARE FRAMEWORK

The whistle analysis process fell into several inter-related parts, mostly as a sequential process, but requiring some transfer of data between different modules in a more unstructured way. The programming language C++ was chosen, which allows a modular approach to be taken with programming whilst still allowing fast execution of the resulting program. A graphical user interface was required so that information from different parts of the program could be displayed simultaneously on the screen. At the start of development, no general windowing package was available for this compiler, and no links were available to enable it to use systems such as Microsoft Windows. For these reasons it was decided to write a simple windowing system that could be used for other software packages written with the compiler.

The windows system was designed around just simple areas of the screen, with additional elements being built on top of that in a hierarchical structure. Objects could contain other objects, and the tools that were made available included variable text boxes, framed boxes, push buttons, toggle buttons, user input boxes (text and/or numeric), pull down menus, and composite windows. Thus, the spectrogram of a whistle could be displayed inside a window on the screen and other information displayed concurrently.

4. SOFTWARE IMPLEMENTATION

4.1 Sound File Data Formats

Several different data formats were needed for storing the whistles at different stages of processing. Initially, the raw signal data was read from a file. This information was stored as a stream of 16-bit numbers, which corresponds to the 'raw' format for many sound file manipulation utilities. Although the RIFF, or 'wav' format is certainly more widely used, these files can easily be converted to 'raw' files.

After loading, the sound signal is converted to a spectrogram using a fast Fourier transform (FFT) routine. Since this format is the one most often used in the analysis routines, a file format was developed for saving the spectrogram information to disk as a '.fft' file.

Once the whistle's time-frequency contour has been extracted from the surrounding spectrogram, it can be encoded as a string of frequency-time-intensity tuples instead of the entire spectrum. This format normally reduces the file size by a factor of 100 (e.g. a 'typical' whistle of 1.2 seconds was stored in an '.fft' file of 185 kb, but a contour encoded '.ctr' file of 1.5 kb).

The next stage of contour processing is to encode it as a series of 'segments' according to general shape. The contour in each segment is modelled by a quadratic equation of the form:

$$y(x) = a_0 + a_1 x + a_2 x^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(1)$$

Since this form very much reduces the data requirements for whistle representation, and also forms the representation used immediately prior to the classification process, a file format for an encoded whistle contour '.ewc' was produced. This format contains the most compact representation of the whistle's contour (e.g. the previous 1.5 kb '.ctr' file is encoded in 344 bytes).

4.2 Viewing the Whistle Signal

The raw whistle signal is read in as a series of amplitudes from the file. This unprocessed information can be viewed in a window on the screen in the standard waveform representation (Figure1). Two slide bars are available for repositioning the centre of the window to any point along the signal, and both the amplitude and time scales can be modified in pre-set increments or to any specific value. The side of the signal window shows scales for both the time (in seconds or sample number) and amplitude (between −1 and +1).
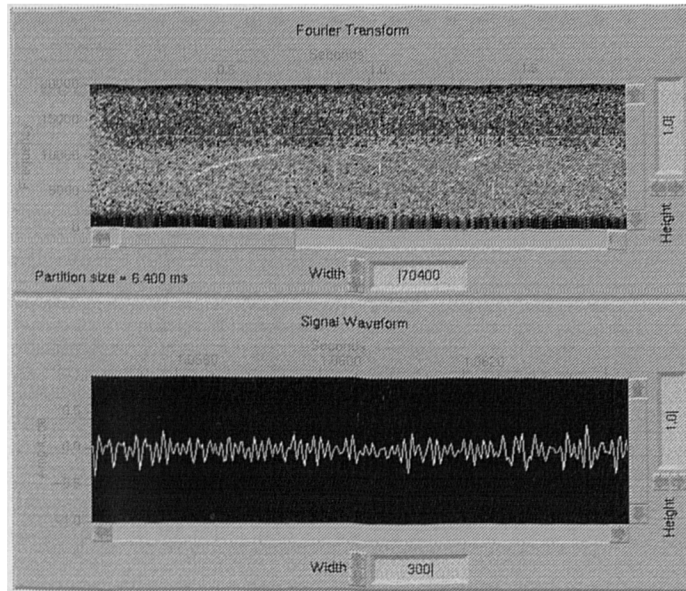
Figure 1: Waveform and FFT windows

Since it is generally quite difficult to separate simultaneous signals from the raw data, the time-amplitude data was passed through a discrete fast Fourier transform routine so that it could be represented in a time-frequency-intensity manner. Tonal signals become distinct from more broadband sounds when this representation is used. A window is used to display the transformed signal that is similar to the one used for the raw signal. The scales in this case are time on the x-axis, and frequency on the y-axis. The intensity can be represented in three different ways using a black to white, a colour-cycling, or simple two-colour palette.

No information on the sample rate at which the file was recorded is stored in the raw signal file. Since this affects both the frequency scale in the FFT window, and the time scales in both FFT and signal windows, a default of 44.1 kHz is used unless the user specifies a different value. Similarly, the time partition for the FFT defaults to 256 sample, giving 128 equally spaced frequency bins displayable on the screen. This can be reduced to increase the time resolution or increased to provide better frequency resolution (each at the expense of the other parameter, of course).

It is possible to save the FFT data once it has been derived from the raw signal. Similarly, one can load the FFT data and recalculate the initial signal from it. The user can supply the format in which he wants to save to the data file from a drop down menu.
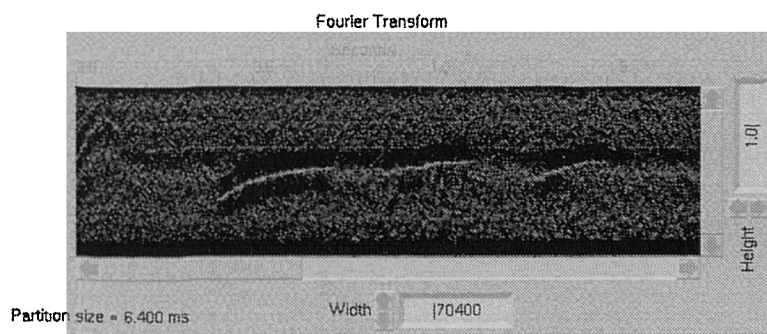
Figure 2: Whistle spectrogram after background noise filtering

4.3 Background Noise Filtering
It is possible for the human eye to determine which parts of the spectrogram consist of background noise and which are the required whistle signals, but any computer algorithm for this task would need to search forward and backward through the spectrogram with quite complex calculations to determine which part was noise. Far easier is to develop an algorithm

that reduces background noise, and then to have a simpler whistle contour following algorithm.

The algorithm of choice for noise reduction leaves a low intensity 'trough' around any tonal components in the spectrogram (Figure 2). The equations employed to achieve this filtering are detailed elsewhere in these proceedings [11].

4.4 Contour Extraction and Encoding

If we visually trace a whistle's contour from a spectrogram, we take account of which way the contour was heading previously when judging where it goes next. This contextual information improves contour following performance considerably over a straight 'ridge-following' technique, and it was included in the contour extraction mechanism used in the software. A 'notional point' was used to trace through the contour, the motion of which was controlled by a number of parameters (Figure 3).

Our notional point moves through the spectrogram with the aim of following the track of highest intensity. Each point on the spectrogram immediately in front of the point exerts a pull proportional to its intensity towards itself. Thus, when no account is made of any previous direction of travel, the notional point becomes the point that produced the strongest pull.
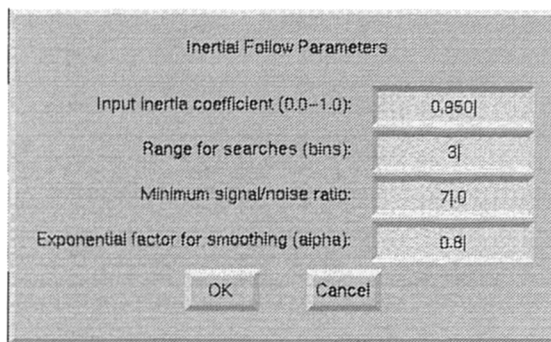
Figure 3: Parameters and typical values for the inertial following algorithm

The 'inertial coefficient' shown in Figure 3 determines how much weight is applied to the previous direction of travel when calculating the new direction the notional point should take. Equation 2 shows the way that the direction of travel is calculated.

$$V'_{xy}(k,t+1) = \alpha_{inertia}V_{xy}(k,t) + (1 - \alpha_{inertia})\left(\frac{\Delta}{|\Delta|}|f'(k,t+1)|^2\right) \dots\dots\dots\dots (2)$$

If the inertial coefficient is set to 1.0, the direction of travel V stays the same as it was for the previous time partition. If it is 0.0, the direction of travel becomes wholly dependent on the direction to the new point (represented by the Δ component) and the intensity at that point. By varying the inertial coefficient between 0.0 and 1.0, it is possible to specify the sensitivity of the tracing algorithm to changes in direction of the contour. The next point on the contour is the point that results in the largest magnitude for the direction of travel vector, V.

The algorithm only searches within a set number of frequency bins of the previous position when searching for further points on the contour. This reduces the time taken to trace the contour, since dolphin whistles very rarely have sudden changes in frequency. However, some whistles may contain rapidly changing frequencies, so the user can alter this search range.

The minimum signal to noise ratio sets the point at which the algorithm is to finish tracing the contour. The same variable is used to determine when a contour begins or restarts after a gap. However, the background noise in a spectrogram sometimes masks the signal for one or two time partitions. In order to overcome this, an exponential smoothing factor was introduced which takes a running average over the previous partition (Equation 3). S in the equation indicates the current spectrum. Values for α in the equation were between 0.0 and 1.0, similarly to the inertial coefficient used in Equation 2. Values near 1.0 produce a long

time average, and those near 0.0 a short one. Spectrum smoothing is used for determining both the start and the end of the contour.

$$S'_{ave} = \alpha S_{ave} + (1-\alpha)S \quad \text{.. (3)}$$

Once the contour has been extracted from the spectrogram, it is encoded in a series of segments. Each of these segments contains areas where the contour is generally rising, falling, flat, or temporarily absent. A segment's contour is then modelled by a quadratic curve (Equation 1), with the origin reset for each segment, allowing them to be compared easily. An extracted whistle contour and the same contour after encoding can be seen in Figure 4. Encoding is carried out automatically in the program with no user intervention required.
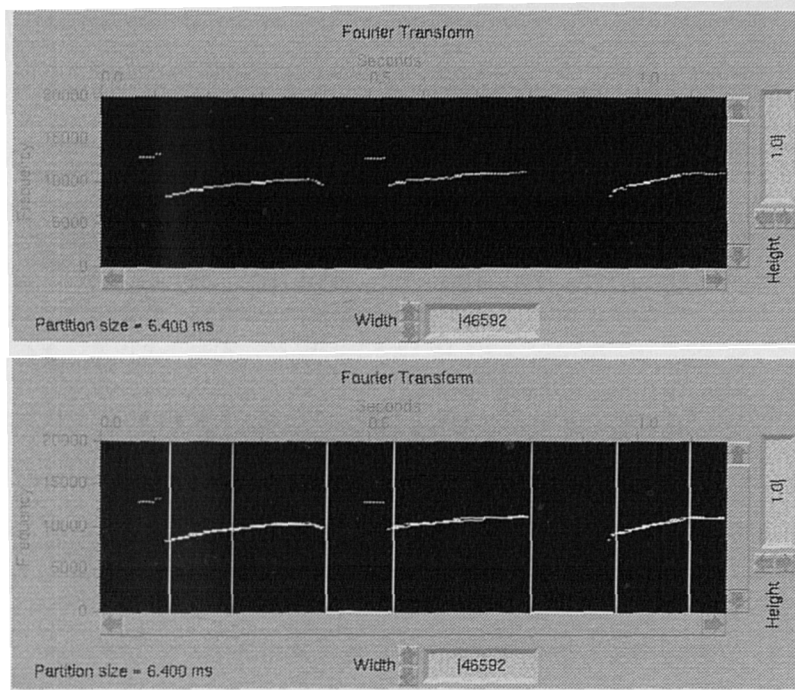
Figure 4: Isolated whistle time-frequency contour before and after encoding

4.5 Classification

Classification is also a predominantly automatic activity from the user's point of view. The software can provide listings of the classes already constructed, and their component whistles, or can find the most probably class for a whistle. When an encoded whistle is to be added to a class, the software calculates the top five most probable classes for it, including the probability it belongs to none of them. The user then has the final decision as to which class to assign the whistle. This intervention can be important if, for example, two classes contain quite similar whistles giving membership probabilities of 27% and 33%, and the probability of a new class is 40%. In this situation, the user might wish to assign the whistle to the class with
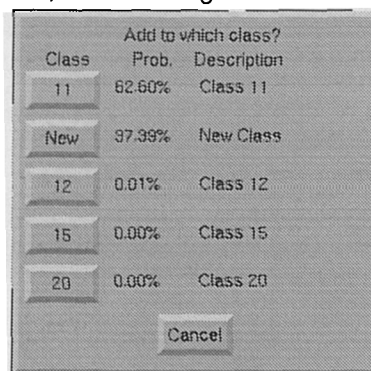

Figure 5: Class assignment prompt, showing top five classes

the 33% probability rather than assigning it to a 'new' class, since the combined probabilities

of the two existing classes is 60%, outweighing the 40% for the 'new' class assignment. An example of a class assignment window for a whistle is shown in Figure 5.

5. FURTHER ANALYSIS

Many recordings contain whistles that do not seem to be associated with group or individual identities. In these cases one cannot examine the probability of one whistle belong to an existing set of classes and determine if that dolphin's group is the same as that previously encountered. Instead information from a number of whistles must be classified for each group, and the similarity in class descriptions used to determine how similar the two sets of whistles are.

	A Classes	B Classes
Group A	5.72	1.90
Group B	2.46	9.56

$\chi^2 = 3.784$, 1 d.f., $p = 1.75\%$

Table 1: Example table of expected whistle distribution into sets of classes.

An example might be for two groups of dolphins to have been recorded: one in the morning and the other in the afternoon. The distinctiveness of the two groups must be assessed. This question can be answered by constructing separate sets of classes for the morning and afternoon groups. The similarity between the groups can be measured by assessing the distribution of the whistles throughout the classes. If the groups are actually the same, then the whistles from the morning group when sorted into classes for the afternoon group whistles should fall into a similar distribution, and vice versa. The software provides the probabilities of a whistle's membership in an existing class or some as yet undiscovered class. So the probabilities can be summed to form a table showing expected number of whistles from the two groups falling into each of the two sets of classes.

In the example shown in Table 1, the sum of membership probabilities for classes formed by whistles in group A was 5.72 for whistles from group A, and 2.46 for whistles in group B. The sum of probabilities for classes from group B whistles was 1.90 for whistles from group A, and 9.56 for whistles in group B. If these two groups consisted of the same dolphins, they should have the same distribution of whistles between the set of classes for A and B. A chi-squared analysis indicates that this is unlikely, and could happen by chance with a probability of 1.75%. Thus, on a comparison of whistle types, there are significant differences at the 5% level between the two groups.

6. CONCLUSIONS

An analysis software tool has been developed that to a large extent can automate the task of pattern recognition of dolphin whistles. Several species of dolphins use identifying whistles, giving this tool a wide applicability to studies of reoccurrence of individual groups. In addition to classifying whistles into individual classes, the results from the software can be used to provide quantitative evidence for the similarity or dissimilarity of two groups.

The analysis techniques presented here still require some human intervention, but a fully automated system that can extract whistle signals from live data and assign it into a class based on its structure requires more research, but is a realistic short-term goal.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] M C CALDWELL & D K CALDWELL, 'Individualized whistle contours in bottlenosed dolphins (*Tursiops truncatus*)', *Nature* (London) **207**, 434–435 (1965).

[2] B MCCOWAN & D REISS, 'Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins (Delphinidae, *Tursiops truncatus*): a re-evaluation of the signature whistle hypothesis', *Ethology* **100**, 194–209 (1995).

[3] M C CALDWELL & D K CALDWELL, 'Statistical evidence for individual signature whistles in Pacific whitesided dolphins, *Lagenorhynchus obliquidens*', *Cetology* **3**, 9 pp. (1971).

[4] M C CALDWELL, D K CALDWELL, & J F MILLER, 'Statistical evidence for individual signature whistles in the spotted dolphin, *Stenella plagiodon*', *Cetology* **16** (1973).

[5] J K FORD & H D FISHER, 'Killer whale (*Orcinus orca*) dialects as an indicator of stocks in British Columbia', *International Whaling Commission Scientific Committee Report* SC/JN81/KW8 (1981).

[6] M E DAHLHEIM, 'Signature information in killer whale calls', *Whalewatcher* **15**(1), 12–13 (1981).

[7] J K FORD & H D FISHER, 'Group specific dialects of killer whales (*Orcinus orca*) in British Columbia', In *Communication and behavior of whales* (R. Payne, ed.), AAAS selected Symposia Series, Westview, Boulder, Colorado, pp. 129–161, (1983).

[8] C R STURTIVANT & S DATTA, 'Techniques to isolate dolphin whistles and other tonal sounds from background noise', *Acoustics Letters* **18**(10), 189–193 (1995).

[9] C R STURTIVANT & S DATTA, 'The isolation from background noise and characterisation of bottlenose dolphin (*Tursiops truncatus*) whistles', *Journal of the Acoustical Society of India* **23**(4), 199–205 (1995).

[10] C R STURTIVANT & S DATTA, 'An acoustic aid for population estimates', *European Research on Cetaceans* — *11* (1997).

[11] C R STURTIVANT & S DATTA, 'Automatic dolphin whistle detection, extraction, encoding, and classification,' (This volume).

Appendix D: Techniques to Isolate Dolphin Whistles and Other Tonal Sounds from Background Noise

C. Sturtivant and S. Datta,
Signal Processing Research Group - Bioacoustics & Sonar,
Electronic & Electrical Engineering Dept.,
Loughborough University,
Leicestershire, United Kingdom

Introduction

During field trials involving a resident population of wild bottlenose dolphins (*Tursiops truncatus*) in the Moray Firth, Scotland, the identity of groups of dolphins was desired for comparison of behaviour. The distance to the animals was often more than 500 metres, which reduced the applicability of the traditional photo-i.d. approach (Wells *et al.* 1980). In order to overcome this problem, the identity of the groups was assessed by the distinctive whistles made by this species, using software to automatically detect the presence of whistles and subsequently extract their frequency/time contours.

It has been noticed by many researchers that several species of dolphin produce distinctive whistles, termed signature whistles, which are associated with individual animals (Caldwell & Caldwell 1965; Caldwell & Caldwell 1971; Caldwell *et al.* 1973). Although this is complicated by dolphins within a group to some extent mimicking each others whistles (Caldwell & Caldwell 1972; Tyack 1986), a group's identity may be determined by its members' signature whistles as the social structure of each group does not change significantly over the one to two week period of each field trial.

Equipment

Recordings of the dolphins' vocalisations were made from a number of sources, including a hydrophone radio telemetry link (sonobuoys) some 500 metres away from the observation base and from directly cabled hydrophones. Each hydrophone was filtered at source to reduce contributions from low frequency sea-state noise below 1 kHz, and was corrected to provide a nominally flat frequency response from this value to at least 20 kHz. The signals were recorded on a four-track Racal Store 4DS intrumentation recorder at $7^1\!/_2$ ips, and included an EBU time-code with a 1/25th second resolution. Details of the times that dolphins were present, their number, group compositions (adults to juveniles), and apparent behaviour were also logged.

These tapes were sampled at 44.1 kHz using a Creative Labs Sound Blaster 16 sound card onto an IBM PC with a 66 MHz 80486DX2 CPU. The tapes were replayed at their original recording speed, allowing frequencies up to 22 kHz to be digitised at a 16-bit resolution.

Methods

Although both tasks required filtering of the Fourier transform of the signal, the task of detecting a whistle's presence was made separate from that of determining the frequency/time contour, as this second task could not be reliably achieved in real time. Tapes were initially replayed and the whistle detection software used to generate a file, listing the times at which potential whistles occurred. The tapes were then replayed a second time and the appropriate parts automatically sampled to disk. This automation replaced the often time-consuming chore of listening to hours of recordings for whistles. Tonal sounds other than whistles were occasionally encountered requiring that the computer's selections be validated.

The second stage was to determine the frequency/time contours of the whistles from the Fourier transform. The signals varied widely in time duration, frequency range, and rate of change of frequency, requiring several algorithm constants to be set manually for each whistle extraction. Significant time savings were achieved over a purely manual determination of the contours, however, and the resulting contours were immediately available for further feature analysis and pattern recognition.

Each of these two stages are dealt with separately below.
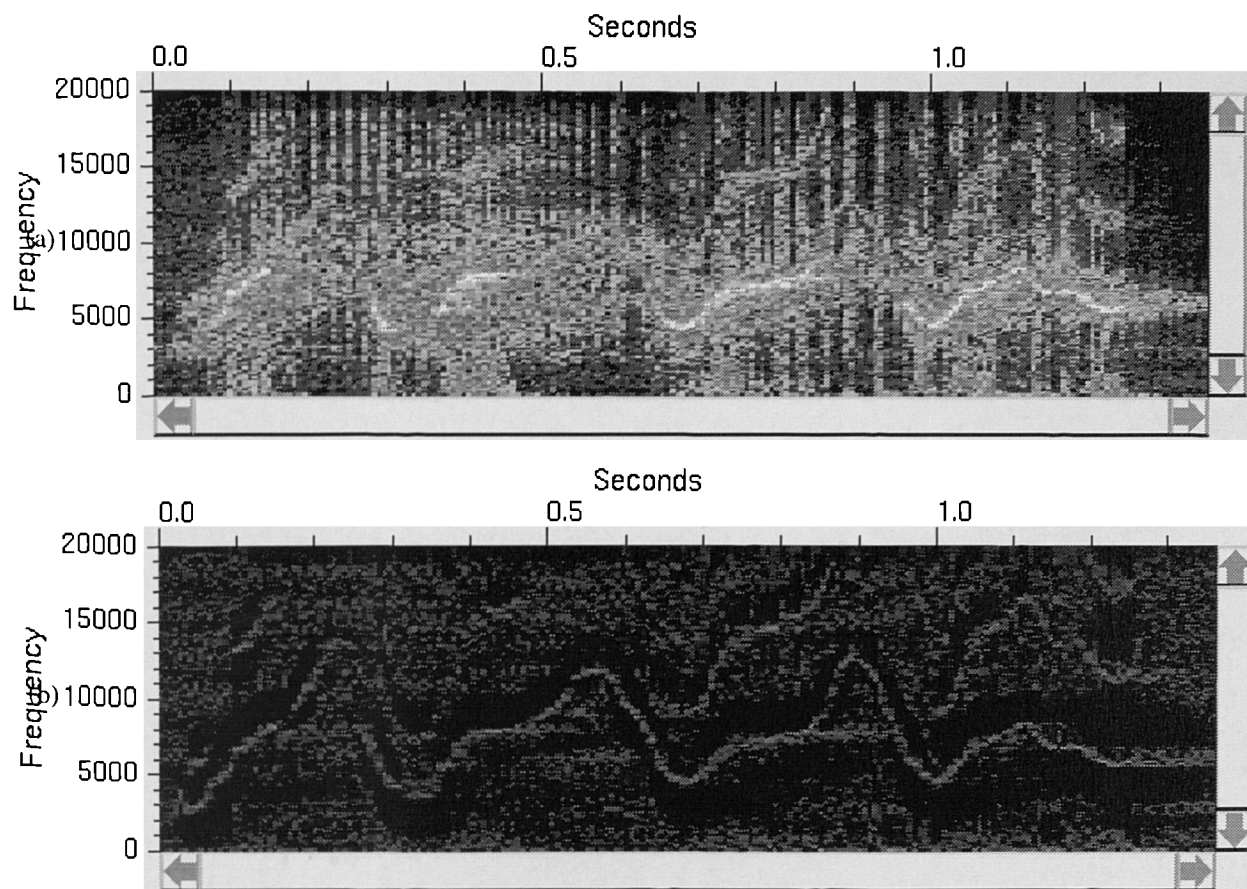
Figure 1 - Spectrogram of a dolphin whistle and clicks: (a) original data; (b) data after filtering

Whistle Detection

Dolphin whistles are characterised by their discrete frequencies and their relatively stable frequency component from one millisecond to the next. A discrete fast Fourier transform (FFT) was carried out on the data to convert the signal into the time/frequency domain. A transform partition size of 256 samples (corresponding to 5.8 ms) produced a good time to frequency resolution for viewing a contour, but for the detection routine a smaller partition size of 32 samples (corresponding to 0.73 ms) was used. This briefer partition time resulted in frequency bins with a width of 689 Hz, so that whistles (which change only slowly in frequency) changed infrequently between adjacent bins.

In order to reduce the contribution to the signal made by impulsive noises (such as echolocation clicks), a filtering technique was used to enhance signals with narrow frequency bandwidths. A typical whistle contour masked by noise from echolocation clicks is shown (Figure 1a). Echolocation clicks generally have a spectrum which changes only slowly with frequency below 22kHz, so an initial method was devised for removal of these clicks whereby the average energy between time partitions was normalised. This calculation is described in Equations 1 and 2.

$$\overline{E}(t) = \frac{1}{N}\sum_{k\,0}^{N-1}|f(k,t)|^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

$\overline{E}(t)$ represents the average energy in time partition t, and $|f(k,t)|^2$ the energy in frequency bin k for time partition t. The normalised energy value E_n was set to a suitable constant value, and the new filtered transform $f'(k,t)$ calculated from Equation 2.

$$f'(k,t) = \sqrt{\frac{E_n}{\overline{E}(t)}}f(k,t) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

However, some remnant of the echolocation click was often present in the filtered transform, as the click spectrum was rarely uniform, especially if the dolphin was not pointing directly at the hydrophone. This prompted a modified algorithm, where the average energy was calculated only from the values in the m surrounding frequency bins (Equation 3).

$$\overline{E}(k,t) = \frac{1}{2m+1}\sum_{l=k-m}^{k+m}|f(l,t)|^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

This new average, now dependent on the frequency bin, was used in the calculation of the filtered transform (Equation 4)

$$f'(k,t) = \sqrt{\frac{E_n}{\overline{E}(k,t)}}f(k,t) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (4)$$

This filtering produced a noise 'trough' around the whistle, as the high energies contained in the whistle frequencies would have a large effect on the average energy calculation for surrounding frequencies (Figure 1b).

In order to remove any background or remaining transient components to the signal, two averages were taken of the resulting filtered signal using an exponential decay (Equation 5). The background noise was assessed with a large value for α giving a "half-life" of several seconds, and the averaged current signal by using smaller values for α giving a half-life of a few milliseconds.

$$x'_{ave} = \alpha x_{ave} + (1-\alpha)x, \quad \text{where } 0 \le \alpha \le 1 \dots\dots\dots\dots\dots\dots\dots\dots (5)$$

The instantaneous (filtered) signal spectrum $f_{inst}(k)$ could then be calculated as the difference between the averaged current noise and the background noise. The mean $\overline{f_{inst}}$ and standard deviation σ_{inst} of the instantaneous signal spectrum was calculated and peaks defined as frequencies whose values exceeded some threshold multiple m_{thresh} of standard deviations above the mean (Equation 6).

$$peak(k) = \begin{cases} 0, \text{ if } & f_{inst}(k) < \overline{f_{inst}} + m_{thresh}\sigma_{inst} \\ 1 \text{ otherwise} & \end{cases} \dots\dots\dots\dots\dots (6)$$

However, using $peak(k)$ to indicate the presence of a whistle was not particularly satisfactory as a high incidence of false alarms occurred, and increasing the value of α for the averaged current signal calculation reduced the algorithm's ability to detect rising or falling whistle contours. Rather than use this method, a list $length(k)$ was kept of frequency bins containing peak values from each time partition, along with the number of time partitions for which the peak had been in existence at that or adjacent frequency bins (Equation 7), with $length(k)$ reset when no peak was indicated. Thus a potential whistle peak could be followed from one frequency bin to the next, and then flagged as a 'detected' whistle in $detect(k)$ when the peak had been in existence for a certain (user-specified) time $length_{thresh}$ (Equation 8). This method significantly reduced the number of false alarms without increasing the number of undetected whistles.

$$length'(k) = \begin{cases} 1, \text{ if } & length(j) = 0 \\ length(j)+1, & length(j) > 0 \end{cases}, \text{ where } j = k-1 \ldots k+1 \quad \text{.......... (7)}$$

$$detect(k) = \begin{cases} 0, \text{ if } length(k) < length_{thresh} \\ 1, \text{ otherwise} \end{cases} \quad \text{...................................... (8)}$$

As the software was not fast enough to carry out whistle detection and store flagged data to disk, a log was made of those times when a detection was made. The times logged began before the initial detection was flagged and ended after the detection was flagged as off, in order that none of the whistle contour was cut off. A second pass of the program over the recording was necessary to sample the data onto the computer's hard disk at the times specified by the log.

Whistle Contour Isolation

The whistle's frequency/time contour is required to allow feature extraction and cluster analysis to be carried out. The detection algorithm reduces the data to be processed to just those parts containing whistles. Determination of the whistle contour is carried out by finding a point on the whistle, and then tracing the whistle backwards and forwards in time from that point. An 'inertial' whistle following technique was used so that overlapping whistle contours could be separated, which weighted following a whistle's pre-existing slope against making sudden changes in direction.

The Fourier transform of the signal was filtered to enhance the whistle's frequency components as detailed previously (Equations 3 and 4), and an initial point $P_{init} = (k_{init}, t_{init})$ was determined on the whistle's contour (Equations 5-8). A point P_f (initially set to P_{init}) was used to keep track of the current place in the spectrogram, and a 'velocity' vector V_{xy} (initially set to zero) was used in conjunction with an inertial component $\alpha_{inertia}$ (set between 0.0 and 1.0) to mediate sudden changes in the contour's direction.

The contour was first traced out for $t < t_{init}$. New velocity values $V'_{xy}(k, t-1)$ were calculated for all points in the previous time partition within a set number of frequency bins, as well as for points immediately above and below P_f according to Equation 9.

$$V'_{xy}(k,t-1) = \alpha_{inertia} V_{xy} - (1 - \alpha_{inertia})(V_{xy} - \frac{\Delta}{|\Delta|}|f'(k,t-1)|^2) \quad \text{.......................... (9)}$$

where Δ is the direction vector from P_f to $(k, t\text{-}1)$.

Thus, the point's 'velocity' V_{xy} is determined by the energy of the whistle contour in the spectrogram, with the inertial factor $\alpha_{inertia}$ determining how much the velocity changes in response to off-vector energy. The next value for P_f was set to the point $(k, t\text{-}1)$ for which $|V'_{xy}(k, t-1)|$ was a maximum. When the signal to noise ratio for the energy at P_f to that in the frequency bins surrounding it had fallen below a preselected threshold value, the start of the whistle was deemed to have been found, and the procedure was repeated for $t > t_{init}$ until the end of the whistle was encountered. After this, a further initial point could be searched for as there might be a break in the whistle contour produced by the dolphin.

A further stipulation for this algorithm was required, as it was possible with low values of $\alpha_{inertia}$ for consecutive values of P_f to oscillate between two high energy points. This was rectified by specifying that no point could be included twice in the same contour. By erasing a contour after it had been traced,

it was possible to follow a concurrent contour, and thus separate the whistles from two simultaneously vocalising dolphins.

Preliminary Results

These techniques were applied to a number of recordings made in the Moray Firth. 101 whistle contours were successfully characterised, including 8 from two animals simultaneously where the contours crossed each other in frequency. A further 15 whistles could be neither characterised using the software nor visually, due to obscuration in background noise.

Acknowledgements

References

Caldwell, M.C. and Caldwell, D.K. (1965). Individualized whistle contours in bottlenosed dolphins, *Tursiops truncatus*. *Nature* **207**, 434-435.

Caldwell, M.C. and Caldwell, D.K. (1971). Statistical evidence for individual signature whistles in the Pacific Whitesided dolphins, *Lagenorhynchus obliquidens*. *Cetology* **3**, 1-9.

Caldwell, M.C. and Caldwell, D.K. (1972). Vocal mimicry in the whistle made by an Atlantic Bottlenosed dolphin. *Cetology* **9**, 1.

Caldwell, M.C., Caldwell, D.K., and Miller, J.F. (1973). Statistical evidence for individual signature whistles in the Spotted dolphin, *Stenella plagiodon*. *Cetology* **16**, 1.

Tyack, P (1986). Whistle repertoires of two bottlenose dolphins, *Tursiops truncatus*: Mimicry of signature whistles? *Behavioural Ecology and Sociobiology* **18**(4), 251-257.

Wells, R.S., Irvine, A.B., & Scott, M.D. (1980). The social ecology of inshore odontocetes. In *Cetacean Behavior: Mechanisms and Functions*, L.M. Herman (ed.), Robert E. Krieger, Florida, pp. 263-317

Appendix E: The Isolation from Background Noise and Characterisation of Bottlenose Dolphin (*Tursiops truncatus*) Whistles[1]

C. Sturtivant and S. Datta,
Electronic and Electrical Engineering, Loughborough University, U.K. LE11 3TU

Abstract

Software routines have been developed to detect and process the tonal whistles produced by dolphins in the wild. Problems in detection due to impulsive noises from echolocation clicks and environmental noises have been overcome using digital filtering techniques, and the time-frequency contours of the whistles extracted from the signals. These algorithms have also enabled the whistles from concurrent whistles to be separated and analysed. Contours of 101 whistles were successfully characterised, including 8 from two simultaneously whistling animals where the contours crossed each other in frequency. A further 15 whistles could neither be characterised visually nor using the algorithms, due to obscuration by background noise.

The data required for input to further pattern recognition routines has been reduced by representation of the whistle contours as a series of syntactic segments, indicated as 'rising', 'falling', or 'flat' in frequency movements, or 'blank' indicating a break in the contour. Within each of these segments, the data points have been approximated to a quadratic curve with a least squares error fit. The use of a segmentation technique, and representation of the segment data with a fitted curve, facilitates a syntactic pattern recognition approach for future analysis of this data whilst also drastically reducing the data requirements for each contour. It is believed that this method better represents the key features used by dolphins to identify each other's individualised whistles.

Introduction

The design of an acoustic reflector to reduce the number of small cetaceans that become entangled in fishing nets has involved field trials with a wild population of bottlenose dolphins (*Tursiops truncatus*) in the Moray Firth, Scotland (Goodson *et al.*, 1994a & 1994b). The identity of the groups of dolphins was required so that changes in behaviour from one encounter with the reflectors to the next could be analysed. However, from the land-based observation point the distance to the animals was often more than 500 metres, rendering the traditional photo-i.d. approach difficult (Wells *et al.*, 1980; Würsig & Würsig, 1977). As the vocalisations from the dolphins in the area of the reflectors were recorded, the whistles from individuals in each group were to be used to determine group identity.

The whistles produced by an individual dolphin frequently appear very similar, and their spectragrams can be distinguished by eye from those of a different animal (Caldwell & Caldwell, 1965). These individualised whistles, termed 'signature' whistles, have been reported for several species of dolphin including the bottlenose dolphin (Caldwell & Caldwell, 1965), the Pacific whitesided dolphin (Caldwell & Caldwell, 1971), and the spotted dolphin (Caldwell *et al.*, 1973). As might be expected, however, dolphins rarely produce one whistle to the exclusion of any others. For instance, dolphins within a group may mimic each others whistles (Caldwell & Caldwell, 1972; Tyack, 1986), one whistle may be repeated a number of times to form a longer one, and its frequency contour may depend on a number of features which may be difficult to observe, such as level of arousal and interactions with other dolphins (Caldwell *et al.*, 1990). Additionally, 'aberrant' whistles may occur which fall in to none of these categories.

If we restrict ourselves to identifying groups of dolphins rather than individual animals (and assume group compositions are fixed for each trial duration), then the problem of mimicry can be ignored, and the group's identity can be determined by comparison of whistles produced by its members with those previously recorded.

Objectives

Recordings of the dolphins' vocalisations were made on a four-track Racal Store 4DS instrumentation recorder at a tape speed of 7½ inches per second, and included an EBU time-code with a 1/25th second resolution. A number of sources were used, including hydrophone radio telemetry links (sonobuoys) deployed approximately 500 metres away from the land-based observation point, and from directly cabled hydrophones. Contributions from sea-state noise were reduced by filtering at source with a high-pass filter with a 3 dB point at 1 kHz, and the signals were corrected to provide a nominally flat frequency response from this value to at least 20 kHz. When dolphins were present, the time, group compositions (number of adults and juveniles), and apparent behaviour were logged.

The main objective was to use these recordings to compare whistles from different interactions and to give a quantitative measure of their similarity, hence providing the probability of the whistles coming from the same dolphin. A more accurate measure of the probability that two groups are the same may be found using a number of such whistles from the groups.

The task was split up in to a number of smaller objectives as follows, with the discrete Fourier transform of the signals used for each stage:

1. Reduction of background noise (e.g. echolocation clicks, wave noise), and enhancement of whistle signals.

2. A detection routine to identify and log those sections of tape in which whistles occurred.

3. Calculation of the whistle's frequency-time contour from the modified Fourier transform.

4. Encoding of the whistle contour in to symbols to aid in matching with similar whistles.

5. A distance measure between whistles for calculation of the probability that two whistles were from the same animal.

The first four of these objectives have been achieved and will be reported in this paper, but the distance measure and probabilities still remain to be completed.

Methods

The recordings were sampled at 44.1 kHz using a Creative Labs Sound Blaster 16 sound card onto an IBM PC with a 66 MHz 80486DX2 CPU. The tapes were replayed at their original recording speed, allowing frequencies up to 22 kHz to be digitised at a 16-bit resolution. The bottlenose dolphin produces few whistles that go above this frequency, so negligible information was lost.

The techniques used for background noise reduction, whistle detection, and frequency-time contour extraction have been reported elsewhere (Sturtivant & Datta, 1995), and only a brief summary of these techniques will be given.
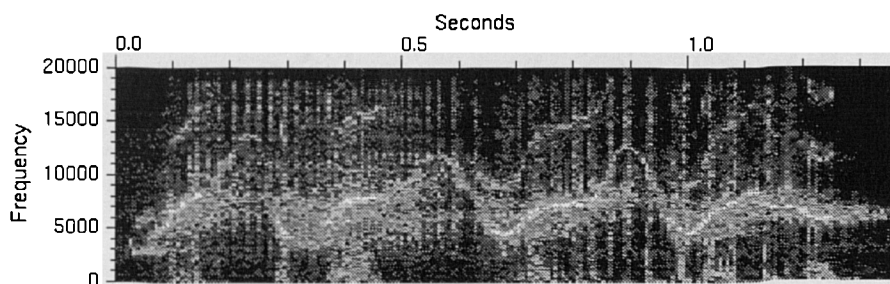

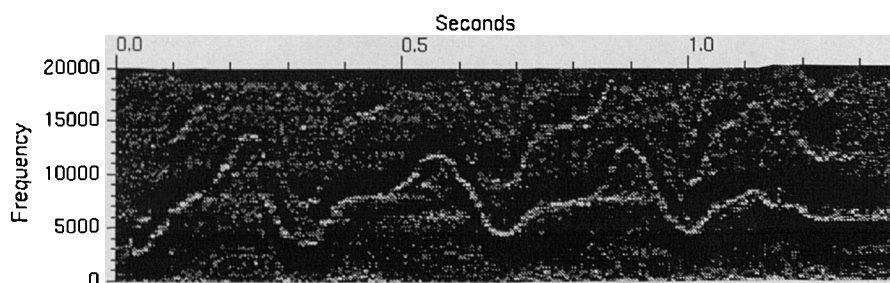*Figure 1. Spectragram showing a dolphin whistle and clicks.*

Figure 2. The same signal after filtering

Background Noise Reduction

Dolphin whistles can be characterised by their discrete frequencies and their relatively stable frequency component from one millisecond to the next. A discrete fast Fourier transform (FFT) was carried out on the data to convert the signal into the time/frequency domain. Observation of the resultant FFT as a 'waterfall' spectrogram showed noise from pops and clicks with a broad spectrum but short time duration, and the required whistles with a narrow spectrum but much longer duration (Figure 1). This fact was used to construct a filter (Equation 1), which de-emphasised components with a wide spectrum and emphasised those with a narrow one.

$$f'(k,t) = \sqrt{\frac{E_n}{\overline{E}(k,t)}} f(k,t) \quad\text{.. (1)}$$

In this equation k represents the frequency bin, t the time partition, and $f(k,t)$ the values of the discrete FFT at that point. E_n is an energy normalisation constant, and $\overline{E}(k,t)$ is the mean energy for the values around the point (k,t) in the range $k \pm m$ where m is a constant. This filter acts in a similar way to an edge-detection routine, resulting in a strong whistle signal and the noise for the surrounding m frequency bins reduced to a very low level (Figure 2).

Whistle Detection

Two moving averages were taken of the filtered signal (using an exponential averaging method) to calculate the long-term background noise, and the short-term 'signal'. The difference between these two averages was then used as the 'current signal'. Peaks in the current signal spectrum were indicated when above a threshold number of standard deviations from the mean, and when these peaks appeared in the same or an adjacent frequency partition for longer than some threshold time duration, then a 'potential whistle' was flagged. Sturtivant and Datta (1995) contains a more detailed description of this algorithm.

Whistle Time-Frequency Contour Extraction

Once a whistle had been detected in a signal, its time-frequency contour was extracted from the FFT. This was done by moving backwards then forwards through the FFT from the point P_{init} at which it was detected. A point P_f (initialised to P_{init}) was used to keep track of the current place in the spectrogram, and a 'velocity' vector V_{xy} (initialised to zero) was used in conjunction with an inertial component $\alpha_{inertia}$ (set between 0.0 and 1.0) to mediate sudden changes in the contour's direction.

The contour was first traced out backwards in time from the point at which it was detected. New velocity values $V'_{xy}(k, t-1)$ were calculated for all points in the previous time partition within a set number of frequency bins, as well as for points immediately above and below P_f according to Equation 2.

$$V'_{xy}(k,t-1) = \alpha_{inertia} V_{xy} + (1 - \alpha_{inertia})(\frac{\Delta}{|\Delta|} |f'(k,t-1)|^2) \quad\text{.......................... (2)}$$

where Δ is the direction vector from P_f to $(k, t-1)$.

Thus, the point's 'velocity' V_{xy} is determined by the energy of the whistle contour in the spectrogram, with the inertial factor $\alpha_{inertia}$ determining how much the velocity changes in response to off-vector energy. The next value for P_f was set to the point $(k, t-1)$ for which $\left|V'_{xy}(k,t-1)\right|$ was a maximum. When the signal to noise ratio for the energy at P_f to that in the frequency bins surrounding it had fallen below a preselected threshold value, the start of the whistle was deemed to have been found, and the procedure was repeated from the point of detection forward. Subsequently, a further initial point could be searched for since a break in the whistle contour was often found to occur.

Attempting to compare one whistle with another by using the simple sequence of time-frequency pairs (extracted by the previous algorithm) clearly would involve comparison of a large amount of data (typically between 0.2 and 3 seconds duration). Observation of the 'shape' of the whistle suggested that a more compact representation could equally well describe the salient characteristics. An algorithm was adopted whereby the whistle was split up in to segments, indicating whether the whistle was 'rising', 'flat', or 'falling' in frequency with time, or 'blank' indicating a break in the contours. The data points contained in each of these segments was represented as a quadratic equation of the familiar form shown in Equation 3.

$$y(x) = a_0 + a_1 x + a_2 x^2 \qquad (3)$$

The values for y and x are normalised over the range of the whistle segment with the origin placed at the first frequency bin in the first time partition of the segment. In general, this quadratic equation would not exactly match all points on the extracted whistle contour, so a least-squares fitting routine was used. The equation was solved by producing a 'design matrix' $\mathbf{A}$, which had the general form shown in Equation 4.

$$A_{ij} = \frac{(x_i)^j}{\sigma_i} \qquad (4)$$

where i ranges over the number of time partitions in the segment, and j from 0 to 2 (from the three powers of x in Equation 3). As the standard deviations for each point were not known, these were all set to 1.0. A column matrix $\mathbf{b}$ was constructed from each of the values y_i for the points in the whistle segment according to Equation 5.

$$b_i = \frac{y_i}{\sigma_i} \qquad (5)$$

where, again, each σ_i was set to 1.0. A precise solution to this problem might then be to find the values of $\mathbf{a}$ (which is the column matrix of the parameters a_i) shown in Equation 6.

$$\mathbf{A} \cdot \mathbf{a} = \mathbf{b} \qquad (6)$$

However, in this case the equation would have no precise solutions, so the least-squares solution using Equation 7 was calculated.

$$\chi^2 = |\mathbf{A} \cdot \mathbf{a} - \mathbf{b}|^2 \qquad (7)$$

The technique of singular value decomposition[2] (SVD) was used to solve this equation. The SVD algorithm takes $\mathbf{A}$, an $M{\times}N$ matrix (where $M{>}N$), and returns three matrices $\mathbf{U}$ ($M{\times}N$ column-orthogonal), $\mathbf{W}$ (diagonal $N{\times}N$), and $\mathbf{V}$ ($N{\times}N$ orthogonal), related to $\mathbf{A}$ by Equation 8.

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T \qquad (8)$$

The inverse of $\mathbf{A}$ is then trivial to calculate, except where the diagonal values in $\mathbf{W}$, w_i, are equal to zero. For zero (or small) values of w_i, the columns of $\mathbf{V}$ form part of the nullspace of the equation, and so their value in the inverse ($1 / w_i$) can be set to zero. With this proviso, $\mathbf{a}$ is calculated using Equation 9.

$$\mathbf{a} = \mathbf{V} \cdot [diag(1 / w_i)] \cdot \mathbf{U}^T \cdot \mathbf{b} \qquad (9)$$

The square of errors at the ends of the segment could then be used to indicate where the segment required to be further subdivided to produce an accurate curve fit..

---

[2] See e.g. *Numerical Recipes: The Art of Scientific Computing*, by W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, Cambridge University Press, Chapter 2.9, for a description and implementation of this algorithm.

Results

The whistle detection routine has been used with recordings of wild dolphins to detect several whistles. A number of these whistles were impossible to detect with the unaided ear, as their frequency contours appeared exclusively beyond the range of human hearing.

Whistles had previously been sampled from recordings made from wild dolphins in the Moray Firth. From these whistles, 101 contours were successfully characterised, including 8 from two simultaneously whistling animals where the contours crossed each other in frequency. A further 15 whistles could be characterised neither visually nor using the algorithms, due to obscuration by background noise.

The determination of the 'segment type' at each point in the whistle contour required a moving average to be kept, as the discrete nature of the frequency bins would otherwise cause too frequent a movement between the 'rise', 'fall', and 'flat' states. The 'amplitude threshold' value to determine when a contour was considered to be 'broken' or 'starting' also had a large effect on the initial number of segments in to which the contour was divided. Despite fine tuning of these parameters both globally and on a whistle by whistle basis, short segments were often indicated due to noise in the original signal which did not fit in to the overall shape of the contour. A typical whistle contour split up in to segments is shown in Figure 3.
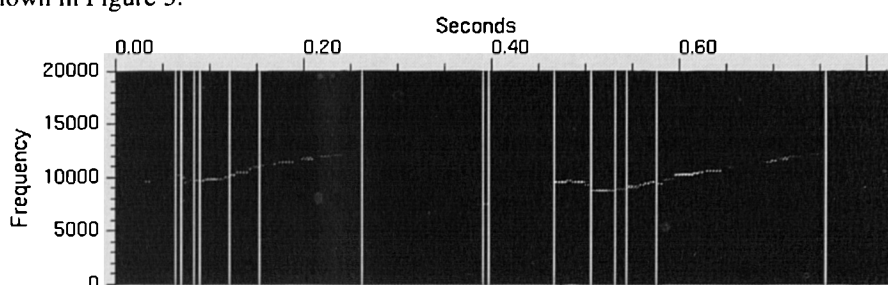

Figure 3. A typical whistle contour after initial segmentation

The noise present in the initial segmentation was reduced by taking small segments, and merging them with larger ones. In order that the least significant segments were removed first, the segment list was scanned to find the briefest segment, and this was merged into its two neighbours. If the two neighbours had the same segment type (as often occurred if the rogue segment was caused by noise), then they were also merged together. This operation was repeated until the smallest segment was larger than some threshold value, chosen to be 1/20th of a second. The 'tidied' segments from the Figure 3 example are shown in Figure 4.
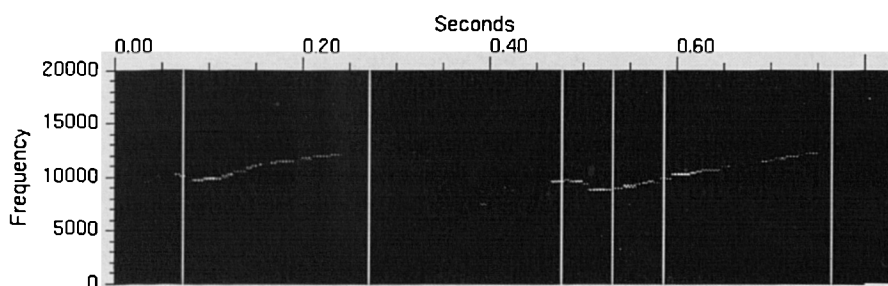

Figure 4. The same whistle after noise segments have been removed.

The SVD least squares error curve fitting was then applied to each segment in the whistle contour, excluding those labelled as 'blank'. These curves generally produced a good fit to the original data (Figures 5 and 6). However, there were occasions when contour segments took on a sigmoid form and so a quadratic equation would not give a good fit to the data.
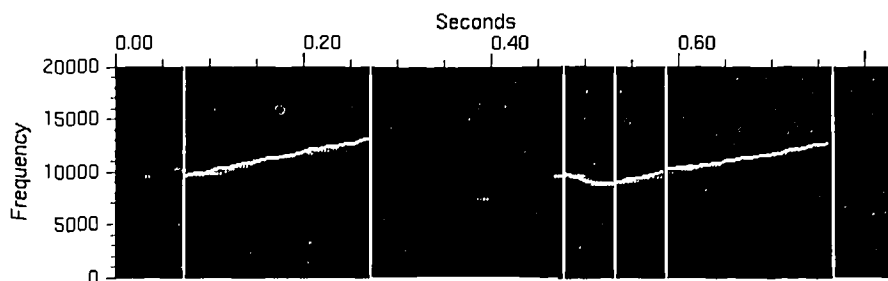
Figure 5. Segments shown after curve-fitting for least squares error
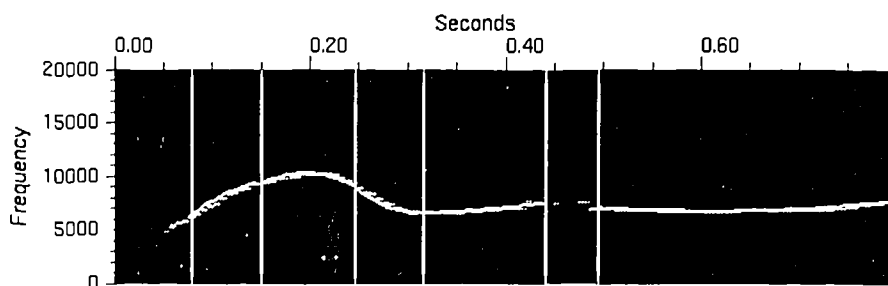

Figure 6. Segments from a second whistle showing curve fitting

A problem with the segmentation of the contours can be seen due to deciding on the segment boundaries solely by a left-to-right scan of the contour. The segment boundaries are thus moved further to the right than would be expected due to the moving average technique used to determine the current slope of the contour (mentioned previously). This effect could be avoided by comparing these segment boundaries with those found using a right-to-left scan, and placing the boundaries at a central point.

Conclusions

Whistles from bottlenose dolphins can be detected in background noise using software algorithms, and their frequency-time contours identified and isolated. The whistle contour can then be represented as a series of segments, each of which is characterised in terms of its shape by the parameters of a quadratic equation. This representation is more appropriate for describing the *shape* of the whistle (which presumably contains the intended information) than is a long list of frequency-time value pairs.

These techniques have been applied to over one hundred whistles from wild bottlenose dolphins in the Moray Firth, Scotland, as well as to oceanarium dolphins where the vocalising individual is known. The contours have been transformed in to a form more suitable for comparison with each other, and more complex pattern recognition techniques, such as dynamic time warping and hidden Markov modelling, can more easily be carried out.

Problems still occur with the placement of segment boundaries further to the right than desired. It is planned to remedy this by comparison with segments calculated from a scan in the opposite direction along the whistle, and taking the mean value of the boundaries for each segment. Additionally, whistles occur where a sigmoid contour shape is contained within one segment, which produces a poor fit with the quadratic equation. Although it is possible to use a cubic equation to calculate a better match in these cases, the comparison of the shapes of cubic curves lends itself less easily to analysis. It is proposed to develop routines which will further subdivide each problem segment so that the sub-segments can each be better matched by the quadratic equation form.

Acknowledgements

References

Caldwell, M.C. and Caldwell, D.K. (1965); Individualized whistle contours in bottlenosed dolphins, *Tursiops truncatus*. *Nature* **207**, 434-435.

Caldwell, M.C. and Caldwell, D.K. (1971); Statistical evidence for individual signature whistles in the Pacific whitesided dolphins, *Lagenorhynchus obliquidens*. *Cetology* **3**, 1-9.

Caldwell, M.C. and Caldwell, D.K. (1972); Vocal mimicry in the whistle made by an Atlantic bottlenosed dolphin. *Cetology* **9**, 1.

Caldwell, M.C., Caldwell, D.K., and Miller, J.F. (1973); Statistical evidence for individual signature whistles in the Spotted dolphin, *Stenella plagiodon*. *Cetology* **16**, 1.

Caldwell, M.C., Caldwell, D.K., and Turner, R.H. (1970); Statistical analysis of the signature whistle of an Atlantic Bottlenosed dolphin with correlation between vocal changes and level arousal. Los Angeles County California Museum of Natural History 8, p. 1

Caldwell, M.C., Caldwell, D.K., and Tyack, P.L. (1990); Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin. In *The Bottlenose Dolphin*, S. Leatherwood and R.R. Reeves (eds.), Academic Press. pp. 199-234.

Goodson, A.D., Klinowska, M., and Bloom, P.R.S. (1994a); Enhancing the acoustic detectability of gillnets. In *Cetaceans And Gillnets, Reports Of The IWC Special Issue 15*. pp. 585-596.

Goodson, A.D., Mayo, R.H., Klinowska, M. and Bloom, P.R.S. (1994b); Field testing passive acoustic devices designed to reduce the entanglement of small cetaceans in fishing gear. In *Cetaceans And Gillnets, Reports Of The IWC Special Issue 15*. pp. 597-606.

Sturtivant, C.R. and Datta, S. (1995); Techniques to isolate dolphin whistles and other tonal sounds from background noise. *Acoustics Letters* (in press).

Tyack, P. (1986); Whistle repertoires of two bottlenose dolphins, *Tursiops truncatus*: Mimicry of signature whistles? *Behavioural Ecology and Sociobiology* **18**(4), 251-257.

Wells, R.S., Irvine, A.B., and Scott, M.D. (1980); The social ecology of inshore odontocetes. In *Cetacean Behavior: Mechanisms and Functions*, L.M. Herman (ed.), Robert E. Krieger, Florida, pp. 263-317.

Würsig, B. and Würsig, M. (1977); The photographic determination of group size, composition, and stability of coastal porpoises (*Tursiops truncatus*). *Science* **198**, pp. 755-756.