Time Frequency Analysis of Dolphin Clicks Reveals Fine Temporal Structure[1]

Judah Jacobson[2] Maryam Saleh[3] Nathan Intrator[4]

Introduction

One of the most effective ways for a dolphin to perceive its environment for the purpose of prey detection is to use its active and passive (listening) sonar capabilities. Acoustic energy propagates in water more efficiently than any other form of energy, so that an active sonar capability is ideal for life in an aquatic environment. Dolphin sound emissions can be classified into two broad categories: narrow-band-frequency-modulated continuous tonal sounds, referred to as whistles, and broadband sonar clicks that have durations between 50 and 200µs (Evans 1973). Passing air back and forth through a set of nasal passages, which creates a narrow beam of sound, first produces the click sound. This sound is then passed through a fat-filled cavity in the head called the melon. In the melon, the sound is culminated so that the dolphin can actively direct the clicks without wasting energy due to scattering the sound. When searching for a target, dolphins usually emit bursts of clicks referred to as a click train. The interval between clicks appears to be a function of the distance from the target. Other factors such as the difficulty of detecting a target, the presence or absence of a target of interest, and the animal's expectation of finding a specific target also affect the click train.

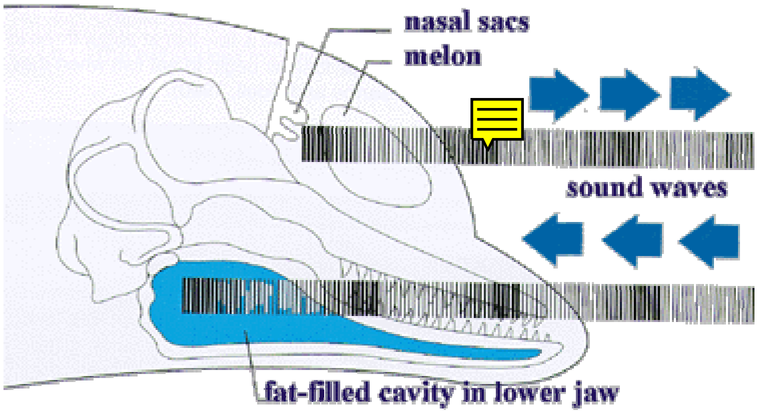


Figure 1. Dolphin Sound Propagation and Reception. Taken from http://www.seaworld.org/bottlenose_dolphin/echodol.html

The ability of dolphins to accurately perceive their environment and to perform difficult recognition and discrimination tasks depends to some degree on the characteristics of their sonar signal. The notion that dolphins make fine adjustments to the spectrum and shape of their click in a pulse train is appealing and has been investigated by several researchers. Bel'kovich and Dubrovsky (1977) argued that any significance attached to changes in spectrum and shape of

---

[1] This work was supported by ONR grant N00014-01-1-0306

[2] Department of Mathematics, Harvard University, One Oxford Street Cambridge, MA 02138.

[3] Engineering Department, Brown University, Providence, RI 02912.

[4] Corresponding author. Institute for Brain and Neural System, Brown University, Providence RI 02912. Phone 401-863-3857, Nathan_Intrator@brown.edu.

dolphin clicks reflects incorrect interpretation of the results or is based on insufficient statistics. The spectral changes observed in their experiment were not consistent among all trials and were attributed to the dolphin turning momentarily away from the target, to internal steering of the beam even when its rostrum was pointed to the target, to fluctuations in the click generation mechanisms, or to lapses in the animal's attention during a trial. Dziedzic and Alcuri (1977) found that the clicks were relatively invariant when the dolphin was more than 4m away from the target. However, when the animal- target distance was less than 4m, they found a spectral spreading of the signal accompanying an increase in the difficulty of the discrimination task. In *The Sonar of Dolphins* (1993)*,* Au argues that these spectral changes were due to changes in the dolphin - hydrophone orientation. He states that *variations in the signal waveform as a function of the angle around a dolphin's head compound the difficulties of obtaining reliable sonar emission data on fine spectral adaptation produced for the purpose of optimizing target discrimination and recognition*. To study spectral changes in a target discrimination task, contact hydrophones should be placed on the head of a dolphin. Dziedzic and Alcuri had the hydrophones placed 1m behind the targets, which led to the dolphin-hydrophone orientation problem. The study of the amplitude-frequency relationship in click trains by Au et al. (1985) and by Moore and Pawloski (1990), gives a general impression of the degree to which a dolphin can control the characteristics of its sonar signals: Dolphins emit broad band signal with two center frequencies and ability to control both frequency and amplitude, mainly as a reaction to environmental noise, with some preference to higher frequencies at higher amplitudes.

In a data that was collected by Moore et al. (1991), three dolphins were trained to perform object detection tasks: two of the dolphins, Tt751F and Tt018M, were trained to respond to the presence of a specific target, while the third, Tt598M, was trained to perform match-to-sample tasks. Each match-to-sample trial consisted of two intervals. In the sample interval, the dolphin inspected a first target via sonar; in the comparison interval, the dolphin was presented with three targets, from which it was required to select a match to the original target. Houser et al. (1999), classified the dolphin clicks that were emitted during the experiments using characteristics of the frequency spectrum. From visual inspection of over 30000 clicks, seven categories of click types were developed based on Boolean characters such as peak frequency, secondary peak frequency if it existed, the locations of both (high or low), and [illegible] number of distinctly bounded regions existing within the 3-dB bandwidth. 54,000 clicks were classified using these categories, automating the classification process using a computer program utilizing the same Boolean characteristics as were used by human inspectors. The spectra were also sorted using an artificial neural network in order to evaluate the choice of click types. Results in general supported the choices, although the network did not perform as well in classifying the less frequent click types. They found that the two dolphins with the same task used different click types with different frequency spreads.

The results of Houser et al. show significant change in click type both between dolphins performing similar tasks and between different intervals of the same task. In addition, they demonstrate some variability in proportion of click type between different positions in the click train. This variation is partially due to the biological system, which generates the signal and presumably, to some extent, to active control of the type of click emitted for different circumstances and at different stages of echolocating an object.

In [illegible] paper, we explore click structure using time/frequency analysis. Using principal component analysis, we demonstrate a an atomic decomposition of the clicks, which can be termed clicklets.

Methodology

There are two main differences between time frequency analysis that is done via a continuous wavelet transform and spectral analysis via Fourier transform. The first]is the uniform uncertainty that is associated with each frequency component that is estimated from the data. This uniform uncertainty is achieved by sampling different frequencies using time windows that are adapted to those frequencies (same number of zero crossing), so that there is same amount of information in each such time/frequency window. In fact using a Gabor wavelet, the boxes defined by Δt x Δf are optimal in terms of their area so that the uncertainty of each is the minimal possible as follows from the uncertainty principle of Heisenberg (Mallat, 1999). The second difference is the actual temporal information for different frequency components. This is demonstrated in Figure 2 that depicts three time/frequency representations of dolphin clicks. The time/frequency resentations were obtained using a continuous wavelet transform via a Gabor wavelet. It is seen that different frequencies do not appear at exactly the same time during the click, thus providing more details than a spectral representation. This additional structure leads to a better characterization of the sequence of clicks during a sequence of consecutive click collection[5] as can be seen in the scatter results below.

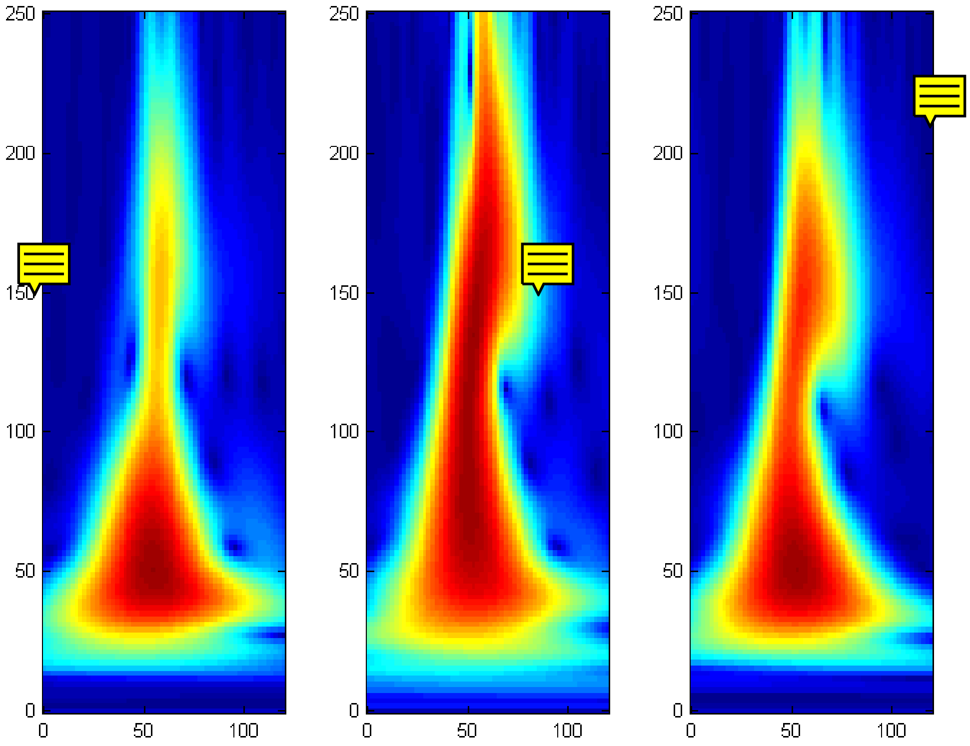


Figure 2. Three typical time/frequency analysis of dolphin clicks. The temporal location of each frequency demonstrates some variability in clicks in terms of the frequency that are used, their duration and the transition from low to medium and high frequency. Note that the middle click shows energy all the way to the top of the scale. In all clicks the high frequency information suggest that there are actually more than one click emitted, or that the click may undergo different paths (through the skull) and thus be expanded in time. Other possibility is some head

---

[5] We call such a sequence a trial although the end of the trial does not reflect the end of the dolphin exploration (unless the number of clicks was smaller than 94).

misalignment between the dolphin and the microphone. We believe that this feature can help in echolocation as the notches structure becomes richer and less ambiguous.

The clicks that we study in this paper were recorded during an experiment in which a dolphin performed a series of match-to-sample trials. In each trial, a sample target was placed in front of the dolphin, which it was allowed to inspect using sonar. After a prescribed delay, three other targets were presented and the dolphin had to identify the target that was used in the previous presentation interval. The individual clicks were recorded at 500 kHz, and 256 sample points were stored for each click. The clicks themselves were very short as can be seen in Figure 2. We thus, used a smaller portion of 128 μsec for click classification, which were represented by 64 time samples and 128 frequency bins.

The time-frequency analysis of the clicks was performed using an analytic wavelet transform with a Gabor wavelet as detailed in (Mallat, 1999). It was compared with a Fourier transform (which does not indicate temporal structure). The clicks were synchronized using a trigger signal so that the time/frequency representation would have minimal variation due to the initial point of the click. We did not include in our analysis clicks that were not synchronized as this usually indicated some transient noise during data collection. Altogether, there were 103 clicks that were not included in the calculation of the principal components (out of a total of 1386 clicks).

In our first attempt to classify the clicks using their wavelet and Fourier transforms, we calculated a correlation distance matrix of the clicks. The high dimensional representation given by the time/frequency analysis (and also by the spectral representation) implies that a robust estimation of any model parameters requires first a reliable dimensionality reduction. In this paper we concentrate on analysis resulting from clicks representation following a dimensionality reduction via principal components analysis.

Principal components analysis (PCA) is the most commonly used method for reducing dimensionality via linear projections. Consequently, this method is the most studied method. PCA is optimal when the high dimensional data has a Gaussian distribution. Some discussion including their limitations can be found in (Duda and Hart, 1973) and some comparison of their properties with higher order moments projections can be found for example, in (Intrator, 1993; Yang and Amari, 1997). While, we do not have any evidence that the high dimensional time/frequency representation of dolphin clicks has a Gaussian distribution, it makes sense to start with it. In particular, it serves our purpose of demonstrating that the time/frequency analysis is relevant to dolphin clicks and provides more information about click structure than spectral representations. The principal components were calculated from all the clicks, which were not rejected due to energy or synchronization problems (as mentioned above). Clicks were taken from the Saline directory. We have used two ways to demonstrate the structure within a click sequence; the first is a clustering of the clicks based on proximity. This is [illegible] via a clustering tree. The distance in y-axis represents the distance between each of the clicks. A natural distance measure is the Euclidian distance between time frequency representations, or between spectral representations. This distance measure was also used between the projections of the representations onto the first six principal components of the data. The following principal components (from 7 and up) still represent interesting structure in time/frequency space; however, they should be ignored, due to their small contribution to the variance of the data as is indicated by the corresponding eigen values[6].. After obtaining the principal components, the high dimensional data are projected onto them to obtain a low-dimensional representation. A tree clustering - dendrogram - of the projections displays much better clustering than the original high dimensional correlation distance matrix (Figures 7,8)

---

[6] We do not present the plot of the eigen values as it is not very informative. It only indicates that over 98% of the variance is explained by the first 8 principal components.

Scatter plots of click sequence provide further information of the dynamics between clicks. They show clusters of sequential clicks, demonstrating some continuity in the click sequence, and show in some cases (so1, co9) some general multi-modal structure of clicks. Since we can not represent structure in a six-dimensional space directly, we look at all two-dimensional projections. Thus, the scatter plots have in their first row projection onto the first principal component (y-axis) vs. projections onto the second to sixth principal component in the x-axis. In the second row the y-axis represents projection onto the second principal component, and the x-axis represent projections onto the third to sixth component. The numbers on the scatter plot represent click sequence, which is also coded in color for convenience.

Results and discussion

We have created two representations to the click signals. The first is a Fourier transform and the second is a Time-Frequency representation. While it is possible to achieve time localization using windowed Fourier transform, this approach is not feasible here as the duration of the click is too short for a robust temporal estimation via windowed Fourier transform.

Figure 3 depicts a sequence of clicks from a single trial in raw data format. For convenience, we remove the axes. The X-axis represents time from 0 to 200μsec. The Y-axis represents linear amplitude scale.

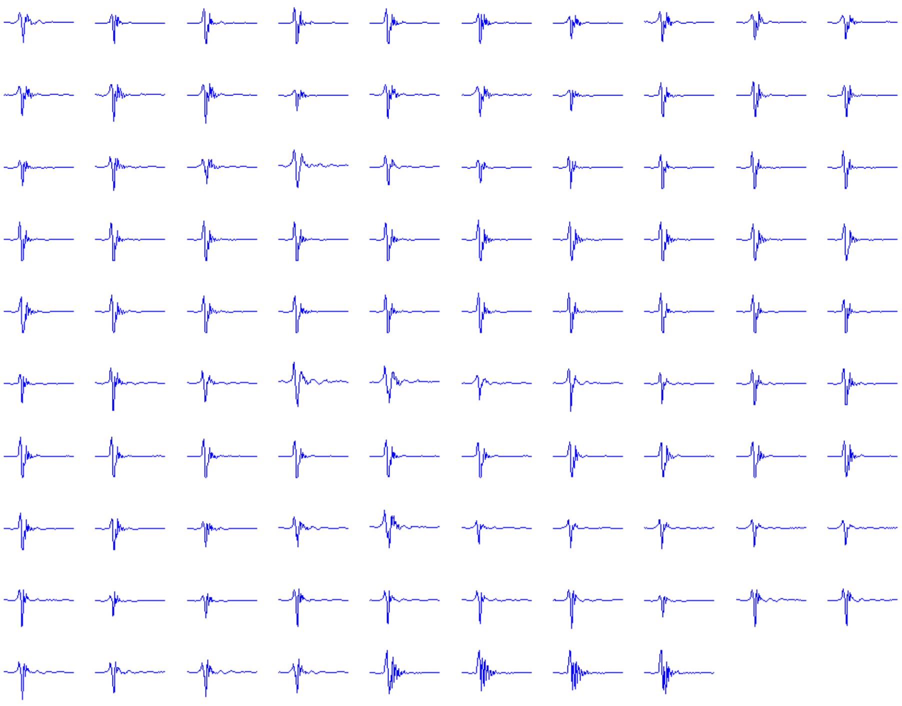


Figure 3. Raw click data. The click itself is very short (about 80 μsec). It is difficult to conclude much about its structure from this representation.

The Fourier representation of these clicks, which is depicted in Figure 4, reveals more structure. This representation was the basis of the analysis of Houser et al. (1999). As was mentioned before, they split the data into seven categories depending on the different spectral structure. Some of these categories can be seen in the Figure, e.g. the sixth click up to the seventeenth look relatively similar and somewhat similar to the clicks in the third row.

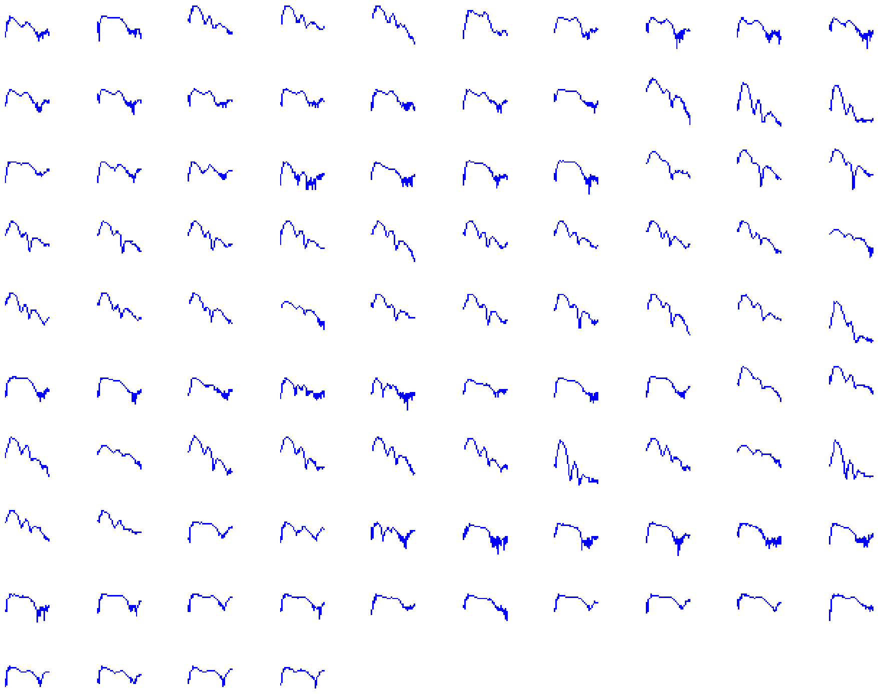


Figure 4. Spectral representation of the sequence of clicks. The amplitude is presented on a logarithmic scale. The X axis spans frequencies from 0 to 250 kHz. Notice some continuity in the sequence, e.g. the fourth row shows very similar spectral representation. The transition from the last few clicks at the third row to the clicks in the fourth row is relatively smooth.

Figure 5 depicts a time frequency representation of the same clicks. There are several apparent findings from this representation: First, it appears that the representation is consistent with the above spectral representation in the sense that it is continuous, namely, the transition from one set of clicks [illegible]e next set occurs at similar locations and those transitions are often smooth, namely occur over a set of clicks and not abruptly. This can be seen for example, in the third row or in the sixth row. Sometimes, there is a chirp-like tilt to the click, e.g., the high frequency tilt in the last few clicks at the end of the first row.

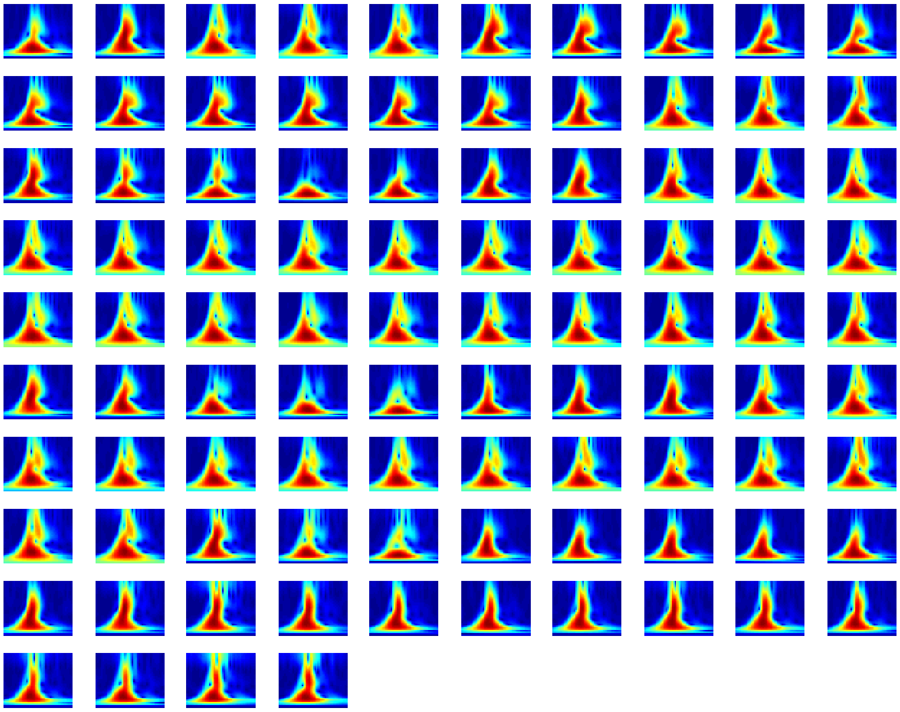


Figure 5. Time/Frequency representation of click trains from a single trial.

The correlation distance matrix for the Fourier transform contains very little clustering, as can be seen in the dendrogram (Figure 6).

The time-frequency results are similar; this is due to the high dimensionality of the data with respect to the number of samples and to the inability to find an appropriate distance measure from the high dimensional representation. A natural dimensionality reduction is via principal components analysis. When performing principal components analysis, it is important to increase the robustness of the dimensionality reduction method, especially when the number of observations is on the order of the number of free parameters that are used in the dimensionality reduction[illegible]re, we calculate a dimensionality reduction in a space that is not larger than the total number of observations used for the reduction. This number is smaller than the dimensionality of the time/frequency representation, and it is thus, more robust to perform the dimensionality reduction in this way (Figures 7 and 8). Following the PCA, we project the data onto the first few components. The number of components is chosen based on the contribution of the component to the overall variance of the data. With these projections, we obtain a representation of the data in fewer dimensions. The results show much clearer clustering, as in Figure 9 with the Fourier and wavelet transforms, respectively. Each shows about 5 different cluster groups; the improved clustering can be seen in the corresponding scatter plots (Figures 10 - 12).

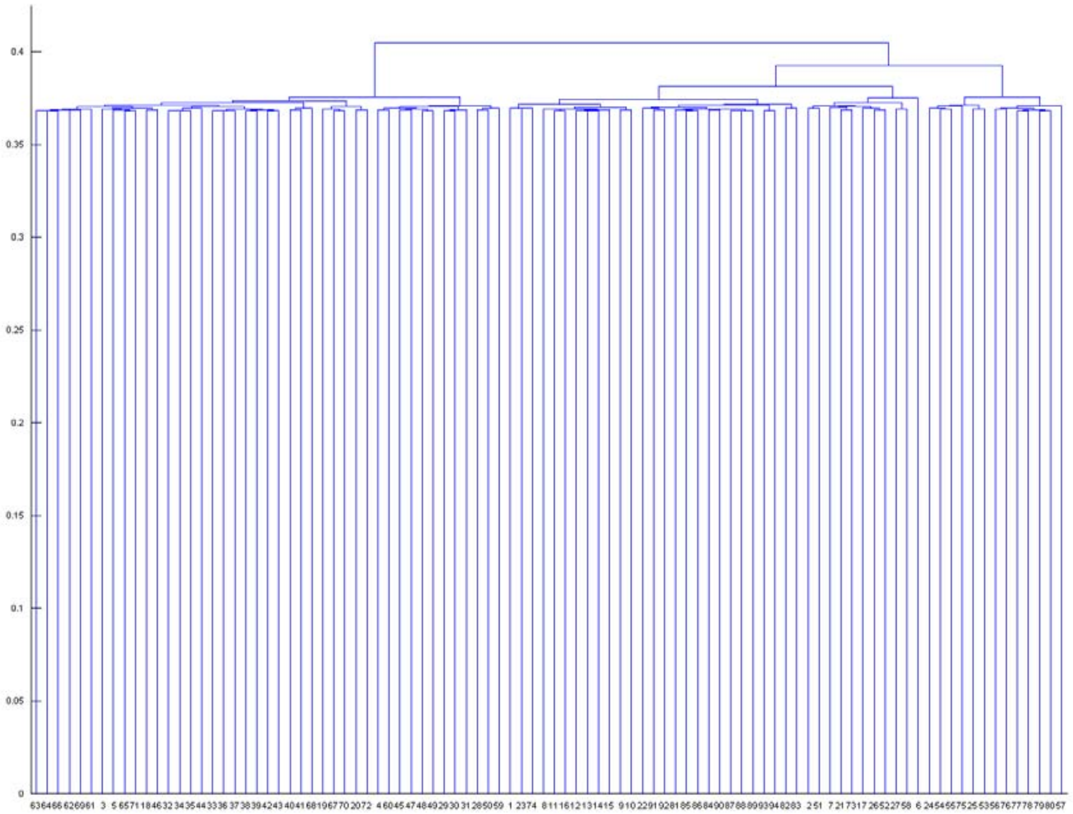


Figure 6. Dendrogram using correlation distance matrix of the Fourier transforms. The dendrogram of the correlation distance matrix for the wavelet transforms looks similar.

Principal components from Fourier and T/F.

The first principal component appears to code for the overall noise level in the system during the collection of the trial. This is observed in the fine changes in background and is thus coded by a strong energy for the background while negating the signal (the blue region). As this noise is likely to related to the environment (and not to the collection system), it is important to correlate [illegible] noise level during a click production with the type of click produced. The other PC's demonstrate a pure structure of the click itself in frequency and time. They demonstrate a certain richness in click production and in particular a slight shift in high frequencies compared to the low frequencies. It is not clear if the shift has to do with some multi-path channel between the dolphin and the hydrophone.

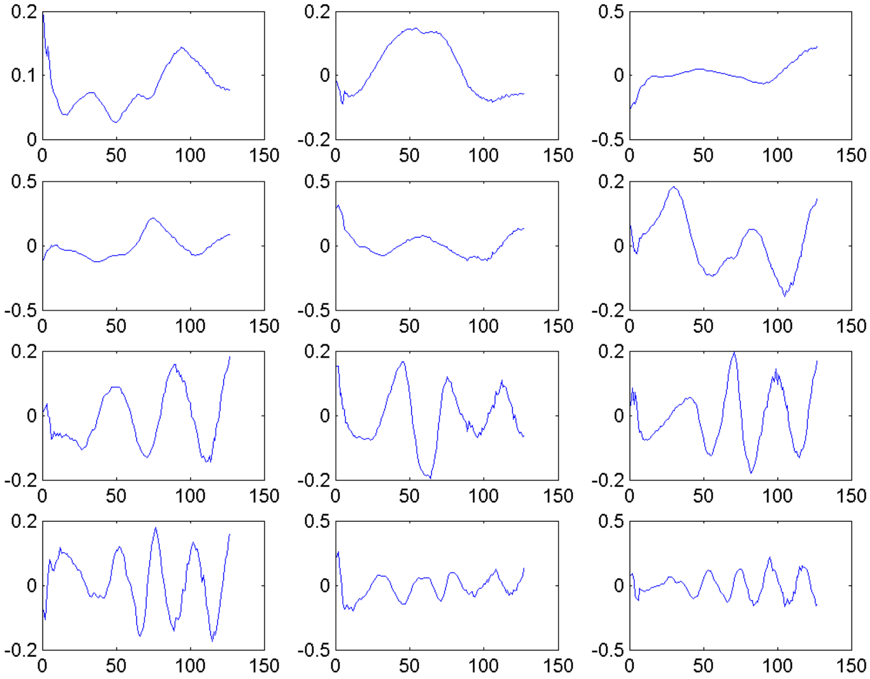


Figure 7. First twelve principal components of Fourier transforms for all files in Saline folder.

The principal components can be interpreted as the atoms of the data representation, as they optimally decompose the data. The principal components obtained from the Fourier representation, demonstrate the same properties described in Houser et al. (1999), namely clicks with a middle and high frequencies, bi-modal frequency clicks etc. It appears that only 5-6 principal components are significant (in terms of contribution to the variance) in the case of the Fourier representation. These components provide further validation to the click types that were found by heuristic methods by Houser et al (1999)

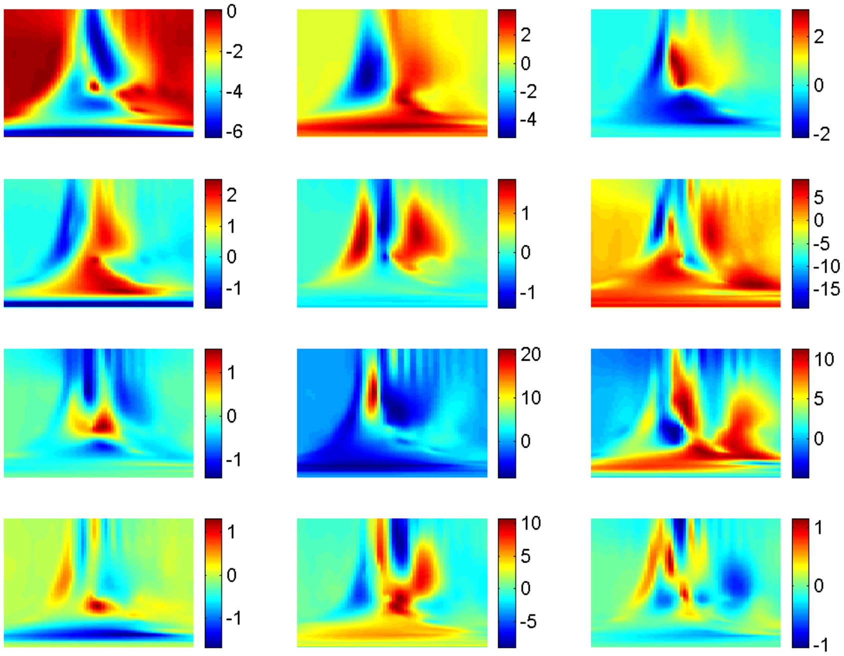
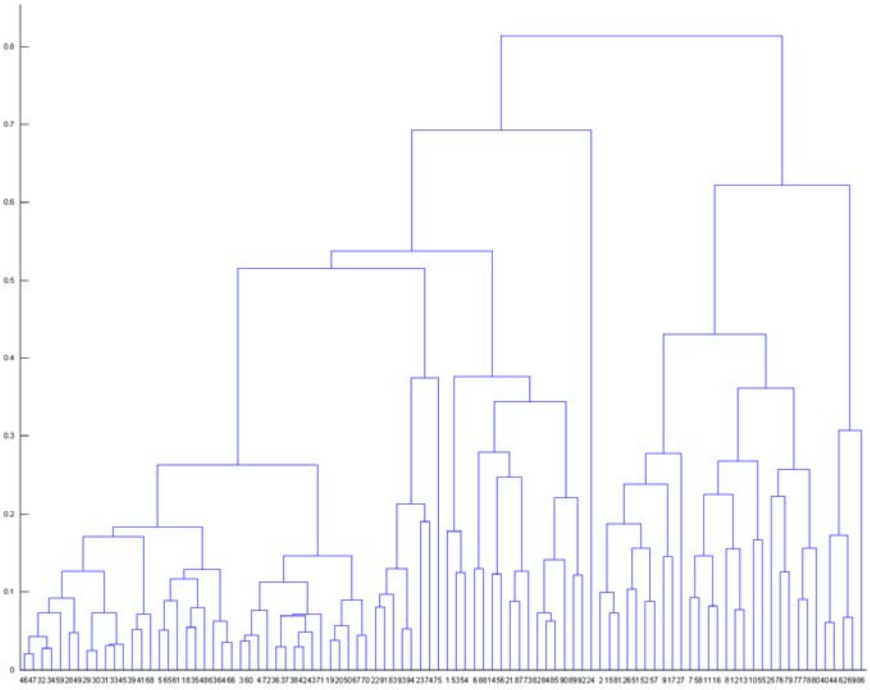


Figure 8. First twelve principal components of time-frequency analysis for all files in Saline folder. The X-axis represents time and spans 128 μSec. The Y-axis represents frequency and spans 0-250 kHz. The amplitude, which is shown in color, represents the amplitude of each coordinate in each component.

Analysis of the principal components obtained from the time/frequency representation indicates several findings: The first component codes the relative ambient noise during the data collection. It can be seen that the values around the locations of the click are very negative, while the locations outside of the click area are negative but close to zero. The next five components are significant in their contribution to the variance. They emphasize sub-click parts and demonstrate the richn[illegible]of the clicks produced by the dolphin. As they decompose the clicks into smaller atoms, we call them clicklets. The sharp boundaries between strong positive values and strong negative values indicate that the localization of the various clicks components was very sharp, namely the production of the click was very consistent. The clicks were obtained from a single experiment of about 20 trials. Signals that did not have strong enough energy or were not synchronized were excluded from the calculation of the principal components, although they later participated in the scatter plots (Fig. 10-12). Synchronization was determined using an automatic triggering mechanism that saved clicks, which produced signals that were above a predefined threshold. The same threshold was also used for the synchronization of the signals.



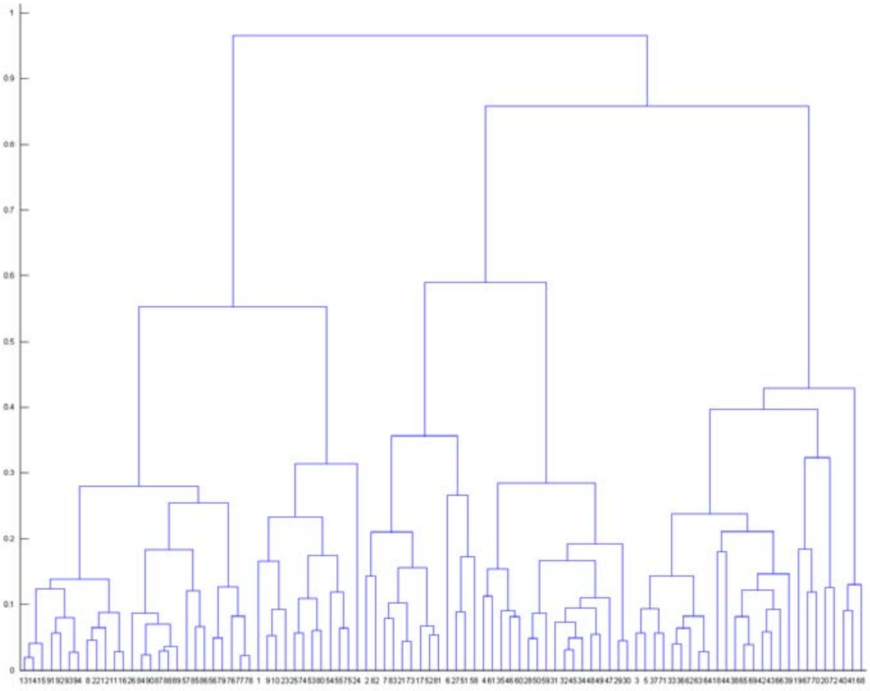


Figure 9. Dendrogram of the projections of the Fourier (top) and Time-Frequency representations (bottom) of one trial onto the first six principal components.

Figure 9 indicates the clusters formed by the clicks when projected on their corresponding principal components. The numbers on the X-axis represent the ordinal number of the click in the sequence of emitted clicks. Thus, numbers that correspond to clicks that are close to each other (according to the tree structure) indicate similar clicks. The similarity is proportional to the total vertical distance between the clicks. There appear to be about 4-5 big clusters, which are approximately the same distance from each other.

A more detailed indication of the clustering is demonstrated by projecting the data on two principal components at a time (Figures 10-12). The scatter plots depict each click by its ordinal number in the sequence as well as using a color scheme that goes from red to blue. In general, the scatter plots resulting from the time/frequency projections appear to contain more compact clusters, indicating that the clicks can not be described by a simple model of a single signal with noise, and is closer to a model of a collection of several prototypical clicks with noise.

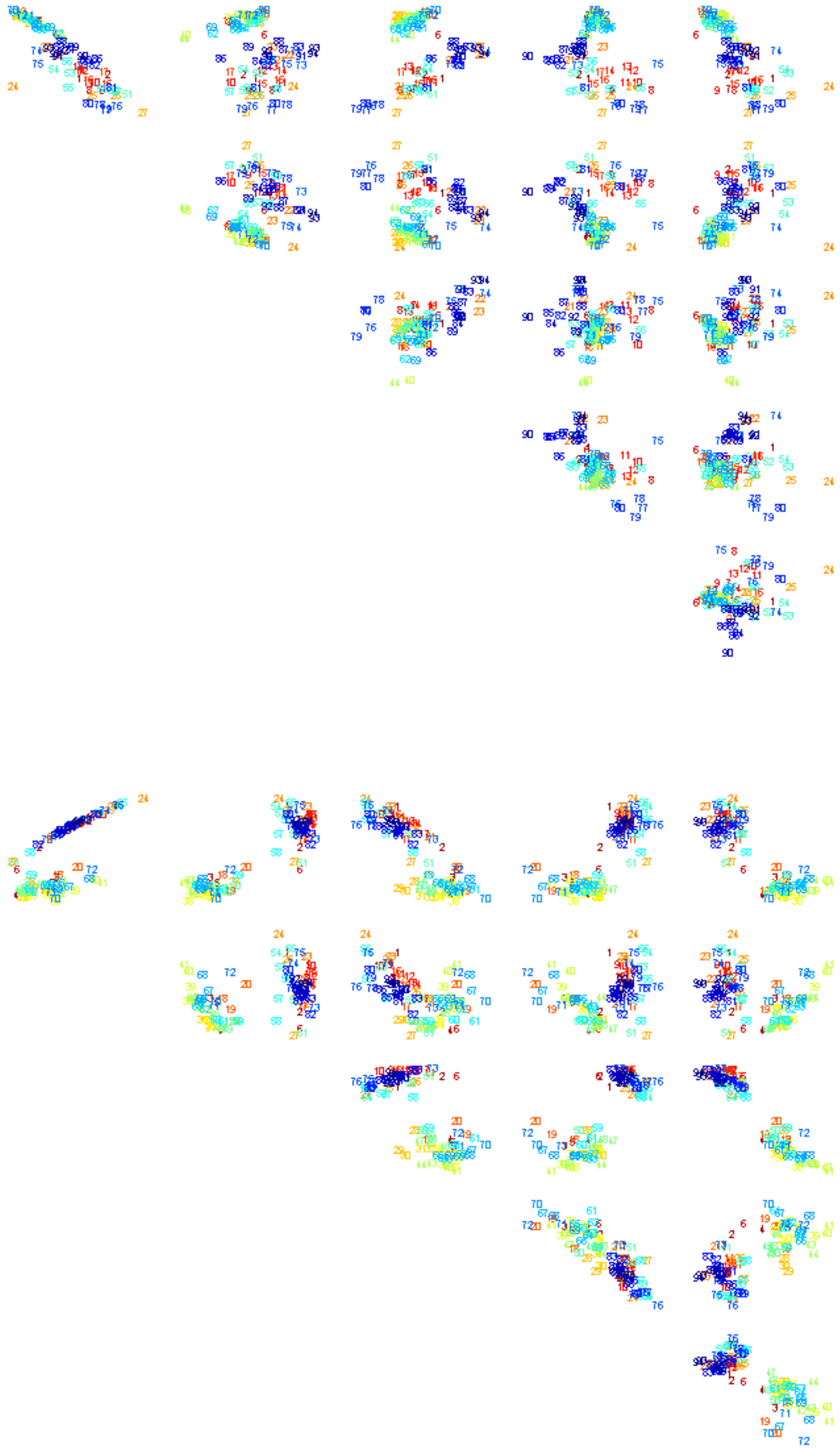


Figure 10. Scatter plots of projections from trial R0606C09 (Saline) onto the principal components. Top: Fourier representation. Bottom: T/F representation.

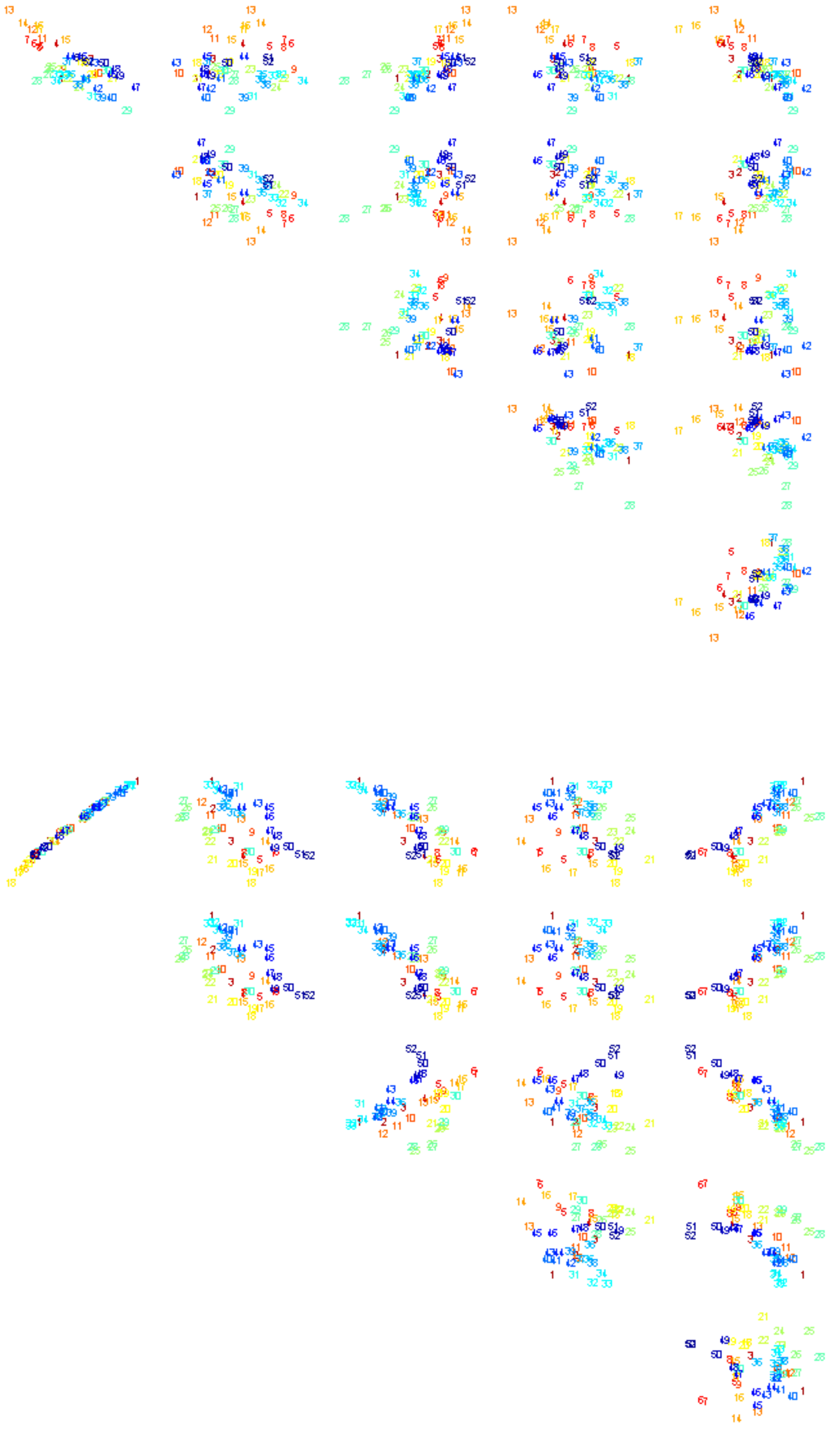


Figure 11. Scatter plots of projections from trial R0606S01 (Saline) onto the principal components. Top: Fourier representation. Bottom: T/F representation.

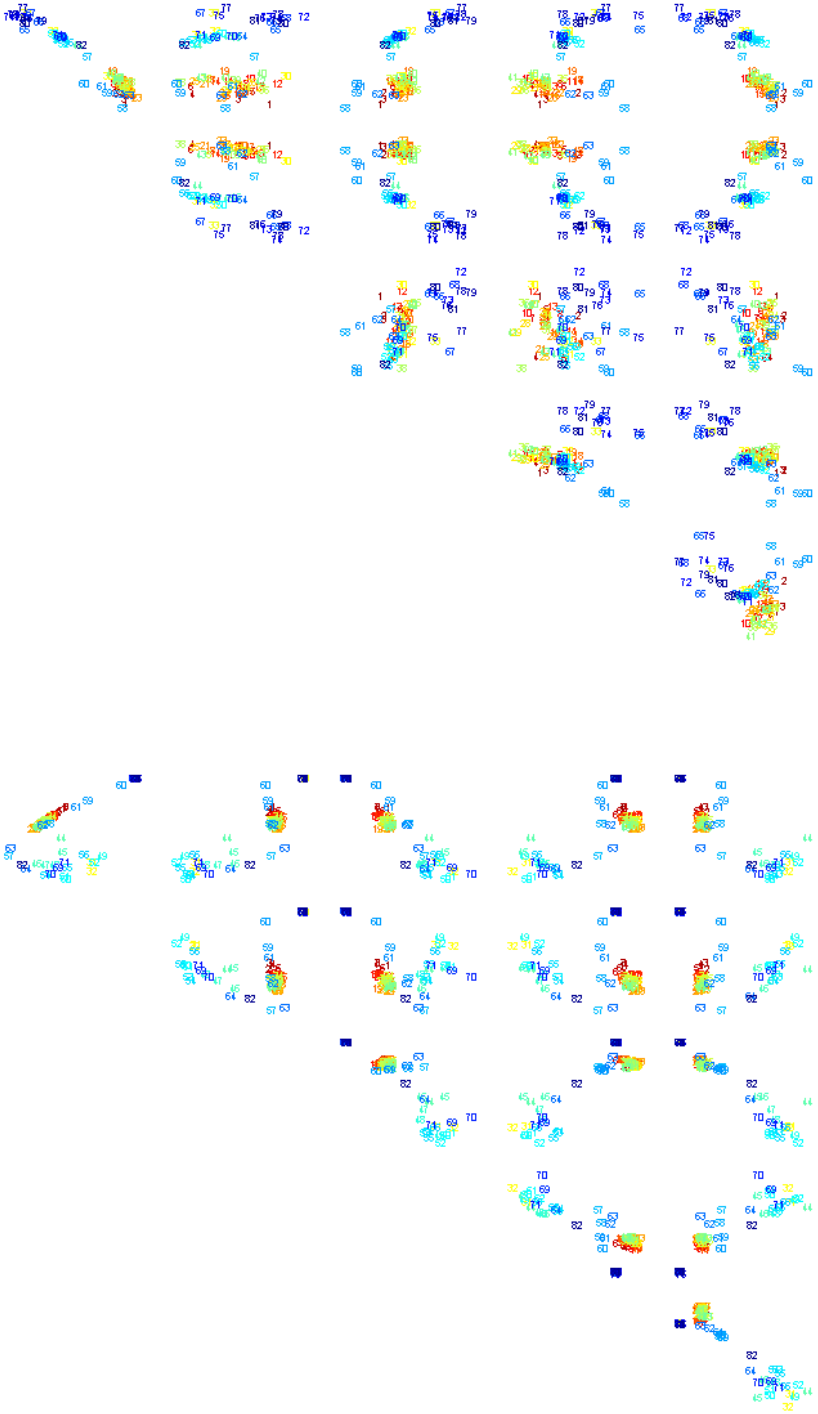


Figure 12 . Scatter plots of projections from trial R0606S16 (Saline) onto the principal components. Top: Fourier representation. Bottom: T/F representation. Note that the rejected clicks are clustered together very closely in T/F but not in Fourier The rejected clicks were numbers 33,65-68, and 72-81.

In fact, as there are clusters of various shapes, it is difficult to consider a Gaussian like noise as a possible explanation to the results. There are some other differences between the Fourier and the time/frequency clusters. The rejected clicks, namely those clicks that were mistakenly recorded usually due to some noise in the triggering mechanism, tend to cluster very far from the clusters of valid clicks, thus, the projections actually validate the fact that these are not proper clicks. This can be seen in Figure 12 (bottom), where the compact dark blue cluster represents incorrect clicks. On the other hand, the same clicks tend to intermix with the projections of valid clicks in the Fourier representation (top) indicating that this representation is not sensitive enough to distinguish between click and non-click signal.

Summary

The aim of this paper was to demonstrate the usefulness of the detailed time frequency representation in analyzing dolphin clicks. We have concentrated on the signal [illegible] by the dolphin as a first step in validating that a time frequency representation that is based on analytic wavelet transform does indeed capture informative structure. We have found that such time/frequency analysis uncovers consistent features which are retained across small sequences of clicks, indicating that this dolphin is performing a certain strategy which includes variation of the temporal order of frequencies, in addition to variation of the frequencies themselves, during target exploration. In particular, this dolphin is generating a collection of signals that cannot be explained by a (single) signal + noise model. Furthermore, there is an interesting cluster structure of the clicks in high dimension, which may indicate biological constraints or signal optimality and need to be further explored.

Acknowledgments

We are indebt to Patrick Moore for providing many helpful comments and insights. We also thank , W. Au, Dorian Houser, Nicola Neretti and Jim Simmons for many helpful discussions.
The Dolphin signal data [Moore et al. 1991] was collected and distributed in the "Rake Database" and is the property of The U.S. Navy, SPAWARSYSCEN, Code D3501B. No further distribution is granted. The Office of Naval Research sponsored the data Collection.

References

Au, W. W. L. (1993). *The Sonar of Dolphins.* New York: Springer.

Bel'kovich, V.M. and Dubrovsky, N.A. (1997). Sensory bases of cetacean orientation. U.S. Joint Publication Research Service JPRSL/7157, May 27, 1977.

Duda R. O. and P. E. Hart (1973). *Pattern Classification and Scene Analys* John Wiley, New York.

Dziedzic, A. and Alcuri, G. (1997). Reconnaissance acoustique des formes et caracteristiques des signaux sonars chez *Tursiops truncatus,*famille des delphinides. C.R. Acad. Sc. Paris 285, Series D, 981-984.

Evans, W.E. ((1973). Echolocation by marine delphinids and one species of fresh-water dolphin. J. Acoust. Soc. Am. 54:191-199.

Houser, D.S., Helweg, D. A., and Moore, P. W. (1999). "Classification of Dolphin Echolocation Clicks by Energy and Frequency Distributions," Journal of Accoustical Society of America, 106, 1579-1585.

Intrator, N. (1993). Combining Exploratory Projection Pursuit And Projection Pursuit Regression With Application To Neural Networks. *Neural Computation*, 5(3), 443—455.Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. Boston: Academic Press.

Moore, P. W. B. and Pawloski, D. (1990). Investigation of the control of echolocation pulses in the dolphin (Tursiops truncates). In: J.A. Thomas and R.A. Kasterlein, eds., *Cetacean Sensory Systems: Field and Laboratory Evidences.* New York : Plenum Press, pp. 305-316.

Moore, P. W. B., Roitblat, H. L., Penner, R. H., and Nachtigall, P. E. (1991). Recognizing successive Dolphin Echoes With an Integrator Gateway Network. *Neural Networks*, 4, 701-709.

Yang H and S. Amari (1997). Adaptive On-Line Learning Algorithms for Blind Separation -- Maximum Entropy and Minimum Mutual Information. *Neural Computation,* 9(7), 1457—1482.