Scale and Translation Invariant Shape and Signal Classification and Detection

William J. Williams
Department of Electrical Engineering and Computer Science
The University of Michigan
and
Quantum Signal LLC
Ann Arbor, MI

ABSTRACT

Highly sophisticated methods for detection and classification of signals and images are available. However, most of these methods are not robust to nonstationary variations such as imposed by Doppler effects or other forms of warping. Fourier methods handle time-shift or frequency shift variations in signals or spatial shifts in images. A number of methods have been developed to overcome these problems. In this paper we discuss some specific approaches that have been motivated by time-frequency analysis. Methodologies developed for images can often be profitably used for fime-frequency analysis as well, since these representations are essentially images. The scale transform introduced by Cohen can join Fourier transforms in providing robust representations. Scale changes are common in many signal and image scenarios. We call the representation which results from appropriate transformations of the object of interest the Scale and Translation Invariant Representation or STIR. The STIR method is summarized and results from machine diagnosis, radar, marine mammal sounds, TMJ sounds, speech and word spotting are discussed. Some of the limitations and variations of the method are discussed to provide a rationale for selection of particular elements of the method.

Keywords: Time-frequency analysis, scale transforms, pattern recognition, radar, machine monitoring, biometrics, word spotting, target identification

1. INTRODUCTION

Much of the early motivation for this work came from attempts to use the sometimes dramatic clarity of time-frequency respresentations of non-conventional signals. These are signals that are not handled well by using conventional Fourier techniques. The Wigner distribution and the spectrogram were well known methods for time frequency analysis,[1, 2] but had their limitations. Starting in the mid to late 1980s a surge of new activity took place and provided a number of new methods for time frequency analysis . This work provided many new possibilities for the representation of signals, but exploitation of the rich detail provided by modern time-frequency representations has received much less attention. One has the impression that further development of time-frequency analysis has slowed considerably in recent time. Time-frequency analysis did not enjoy the hype and quick attention as did wavelet based methods. However, steady progress in using these methods for a variety of applications has continued and perhaps accelerated.

One of the desirable characteristics of a well designed time-frequency representation is scale covariance. Many time-frequency distributions (TFDs) exhibit time-shift and frequency-shift covariance, but only particular attention to design of the TFD also produces scale covariance. Fortunately, TFDs such as the RID easily achieve this goal. Once covariance for these several types of variation is achieved it is possible to produce invariant representations such as STIR. One way that this can be achieved is through application of the Fourier and scale transforms. a brief summary of the methods sets the stage for their application.

williams@quantumsignal.com

Advanced Signal Processing Algorithms, Architectures, and Implementations XIII, edited by Franklin T. Luk, Proceedings of SPIE Vol. 5205 (SPIE,Bellingham, WA, 2003) · 0277-786X/03/$15 · doi: 10.1117/12.513902

2. ANALYSIS TOOLS

The mathematical tools used in time-frequency analysis and scale transforming are briefly covered in this Section.

2.1. Time-Frequency Representations

The basic equation for Cohen's Class of Distributions may be expressed as

$$C_z(t,\omega;\phi) = \int\int \phi(\theta,\tau) A_z(\theta,\tau) e^{-j(t\theta+\omega\tau)} d\theta\, d\tau \tag{1}$$

Here, A_z is the ambiguity function and ϕ is the kernel. The ambiguity function is essentially the inverse Fourier transform of the local autocorrelation, $R(t,\tau)$.

$$A_z(\theta,\tau) = F_t^{-1}[R_z(t,\tau)] \tag{2}$$

where $R(t,\tau) = z(t+\tau/2)z^*(t-\tau/2)$. One may produce any number of TFDs by choosing the shape of the kernel, $\phi(\theta,\tau)$. However, scale covariance of the TFD is achieved by the simple step of making $\phi(\theta,\tau) = \phi(\theta\tau)$, with the latter being termed a product kernel. Other characteristics of the kernel such as reduced interference can still be easily accommodated. Since members of Cohen's Class with time-invariant kernels automatically exhibit time and frequency shift covariance, imposing the product kernel further guarantees scale covariance. We note that wavelet transforms feature time-shift and scale covariance, but not frequency shift covariance. Scale covariance in TFDs has a special form. Since $F[z(at)] = Z(\omega/a)/a$, it is natural for the time frequency result to be $C_z(at,\omega/a)/a$. That is, the TFD appropriately expands/contracts in the time/frequency direction just as does the time and frequency forms of the original signal. This makes it easy to construct an invariant representation.

2.2. The Scale Transform

The scale transform has been described by Cohen[3] to be:

$$D(c) = \frac{1}{\sqrt{2\pi}} \int_0^\infty x(t) \frac{e^{-jc\ln t}}{\sqrt{t}} dt \tag{3}$$

The scale transform is of interest to this paper because it is able to remove the effect of scale in images. There is an analogy to the Fourier transform. The Fourier transform of a signal, $x(t)$ and the Fourier transform of a shifted version of that signal, $x(t-t_o)$ differ only by a phase factor.

$$F[x(t-t_o)] = X_o(\omega) = X(\omega)e^{-j\omega t_o} \tag{4}$$

so that

$$|X(\omega)| = |X_o(\omega)| \tag{5}$$

In a like manner, the scale transform of $\sqrt{a}x(at)$ differs from the scale transform of $x(t)$ only by a phase factor, so that the magnitudes of the scale transform of $x(t)$ and $\sqrt{a}x(at)$ are identical.

$$|D(c)| = |D_a(c)| \tag{6}$$

The phase factor carries the scale value (size) and may be discarded. We have developed discrete forms of the scale transform[4] which can be computed efficiently. One might question the use of the scale transform rather than the more well-known Mellin transform. There are several reasons for using the scale transform. One reason is its easily understood inverse, in contrast to a complicated inverse for the Mellin transform. A second reason is the relationship of scale to wavelet concepts and the insights it brings in this light. The STIR method combines the scale transform and matrix methods to provide a novel means of signal classification.[5]

3. COMBINING METHODS

The first step in the STIR process is to start with a signal or image. If the object of interest is a signal, then a TFD of the signal is computed. It is then possible that a component in the TFD that is of interest is displaced in time and frequency from the ideal template available. It may be possible that the target component is scaled in time and frequency. The following steps are generally taken.

- The TFD is autocorrelated (or FFTed) along time
- The result may or may not be autocorrelated (or FFTed) along frequency

If FFTed, the magnitude of the result is obtained as well. The new object is now invariant to time shift (and frequency shift if the operation is applied in that direction). Whatever the case, the new object now provides an unequivocable zero. This is required for subsequent application of the scale transform. The scale transform is a nonstationary transform and true zero must be the starting point.

One can stop short of the scale transform and use the Translation Invariant Representation (TIR) if scale effects are not present. Also, one may only want to obtain translation invariance over the time dimension if frequency shift is an important discriminator of signal types. If scale invariance is desired a 2D scale transform is carried out on the TIR result and its magnitude is retained. This is the Scale and Translation Invariant Representation.

In some cases, one may wish to apply the method on a one dimensional object such as a signal or spectrum. The FFT spectrum is already time shift invariant, so the STIR is obtained by simply scale transforming it.

If the 2D object of interest is an image, e.g. a face, the same process can be carried out, but the TFD step is not needed. Such images do not enjoy the reciprocal scaling characteristics of TFDs in time and frequency. This may be good or bad depending on the situation. For example, a long face could be mistaken for a round face without some indication of coupling in the x and y directions. These ideas will be illustrated by practical examples later in this paper.

4. DETECTION AND CLASSIFICATION

The process does not stop with the STIR result. A detector or classifier must be designed to take advantage of the attributes of the STIRs. Several methods have been investigated and each has its merits. The preferred method in many cases is the Zero Subspace Method. The Zero Subspace or ZSS is derived from STIR vectors via the Singular Value Decomposition (SVD) method. The SVD is s a method for decomposing a set of vectors into an idealized set of bases such that there is an ordered ability to account for variance according to indexing of the basis vectors. Fortunately, STIR vectors can be formed from the STIR matrices by simply reshaping the matrices into vectors either by concatenating the rows or the columns. Often, there is symmetry which can be used to reduce the number of elements in these vectors.

Suppose that a matrix A is formed by stacking STIR vectors from a set of objects from a given class. Assume that the rows are these vectors. Then, the SVD result can be obtained as

$$USV' = A \tag{7}$$

where the columns of V are bases for the rows of A and S is a diagonal matrix of singular values with $S(1,1)$ being the largest singular value and with monotonically decreasing singular values down the diagonal. If the STIR vectors are of length M and there are exactly M STIR vectors, the columns of V form a complete basis for the STIR vectors. Every STIR vector can be recovered using weighted sums of the basis vectors. If the number of STIR vectors is greater than M, then the set of STIR vectors can be recovered with the lowest mse using weighted sums of the basis vectors. This is the usual case. One wishes for an economy of representation of the rows of A. However, if the number of STIR vectors is less than M, say N, V can be computed to be $M \times M$. The firs N basis vectors are a complete basis for A and hence the STIR vectors. However, there are $M - N$ extra vectors that can be harvested from V and these vectors are orthogonal to all of the STIR vectors.

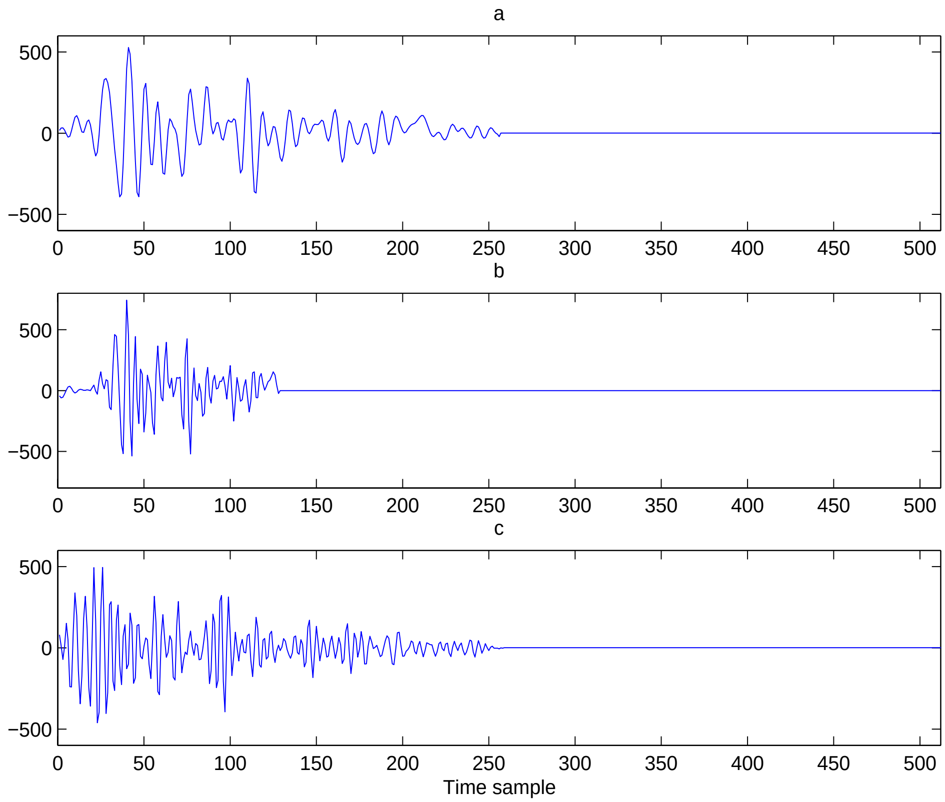


Figure 1. Example of a. original, b. scaled and c. frequency shifted click

We call the space defined by these extra vectors the ZSS. It is to be distinguished from the noise subspace. Suppose that the STIR vectors can be completely represented by 3 basis vectors. The space so defined is the signal subspace. The rest of the first N vectors form the noise subspace and this subspace is orthogonal to the signal subspace. The ZSS is beyond the pale and is orthogonal to both! However, any STIR vector not in the original SVD will likely not be orthogonal to the ZSS. Careful design can make this a powerful tool for detection and classification. Any STIR vector closely related to the STIRs used to form the ZSS will not be completely orthogonal to the ZSS. However, it will likely be more orthogonal than some randomly chosen STIR from another class of objects. Surprisingly, this only works well when all of the STIR vectors are fairly similar.

In practice, the sum of the magnitudes of the challenging STIR ZSS projections is used as a distance measure. Distances for STIRs in the class are less than for STIRs outside the class.

Alternative methods have been examined and sometimes perform equivalently. These are moment based methods and k-nearest neighbor methods. The moment based methods use various moments derived from the image form of the object. One can achieve several types of invariance, including scale and rotation invariance using moment based methods. Some of these alternative methods will be discussed later in the paper.

5. EXAMPLES

An example of the methods is useful. The signal chosen is a single dolphin click. This signal has been widely used as a test signal for time-frequency methods due to its interesting structure. The original click and two transformations are used in the following illustrations. One transformation involves a 2 : 1 time compression of the click along with an energy renormalization. The second transformation involves a frequency shift of the click. These clicks are shown in Fig. 1. The amplitude spectra are shown in Fig. 2. It is clear that these spectra are quite different as presented and would not seem to have originated from the same source.

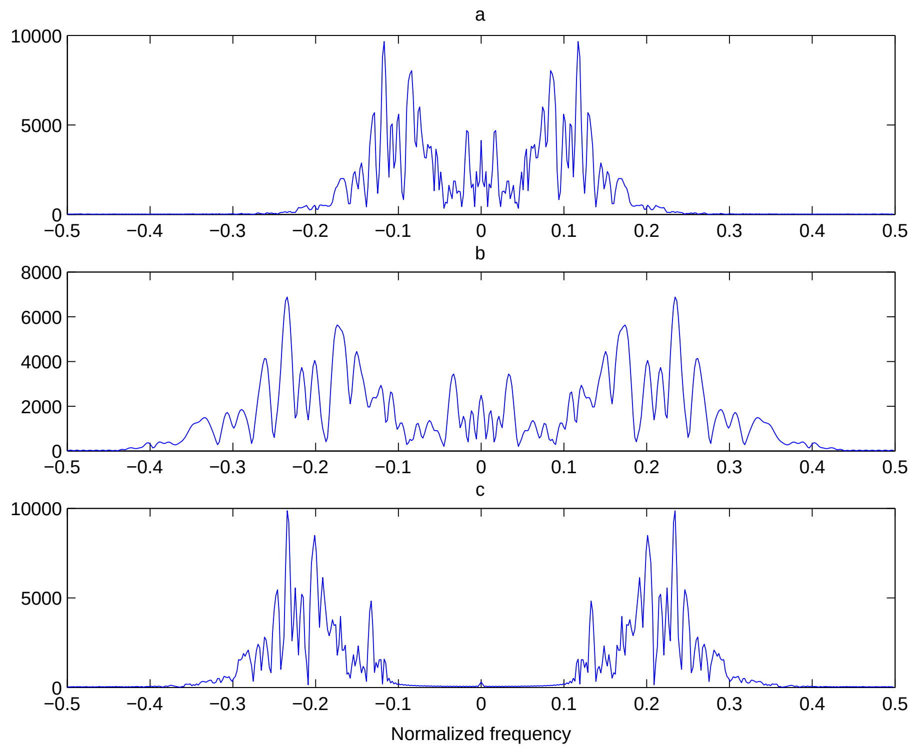


Figure 2. Spectra of a. original, b. scaled and c. frequency shifted click

The next results provide some hope. The binomial TFD is obtained for each version of the click. Fig. 3 compares the TFDs of the original and frequency shifted clicks. The TFD signatures are nearly identical, only differing by the frequency shift and a small time difference. A comparison between the TFD of the original click and the scaled click is provided in Fig. 4 Notice that the time and frequency axes have been adjusted for the scaled TFD such that the signatures are nearly identical except for a small time and frequency shift. This serves to illustrate the point that the binomial TFD is covariant to time-shift, frequency-shift and scale.

If representations are covariant to some known effect and an easy transformation exists it is possible to provide *invariant* forms. It is well known that the magnitude of the Fourier transform is invariant to time shift, for example. The magnitude of the scale transform is invariant to scale. Two steps provide invariant representations of the three click TFDs. First, a 2D FFT (or inverse FFT) of the TFDs is obtained and the magnitude is retained in each case. The magnitudes of the TFDs of the three clicks are shown in Fig. 5. One can see that the signatures are very similar, but the scaled click signature is expanded along the x-axis and compressed along the y-axis. The axes are not labeled, but the y-axis would logically be time and the x-axis frequency in terms of dimensions. These are essentially ambiguity function magnitudes. Finally, a 2D discrete scale transform of each of these results is carried out and the magnitude is retained. These results are shown in Fig. 6. This is the result for one quadrant of the matrix. Two of each of the four quadrants are the same due to symmetry. In practice, only two quadrants are required to capture all of the information. Shown is one component of two of the full STIR. One can see that these results are very similar. The (usual) final step is to reshape each of the two unique results into vectors and combine them to form a STIR vector. On occasion, one can provide a simpler result by just scale transforming the spectra to provide a one dimensional STIR result. Generally, the STIR vectors representing a range of variations of the signal are stacked in rows to form a matrix which is decomposed using the SVD method. Feature vectors for classification and detection are then selected from the SVD results.

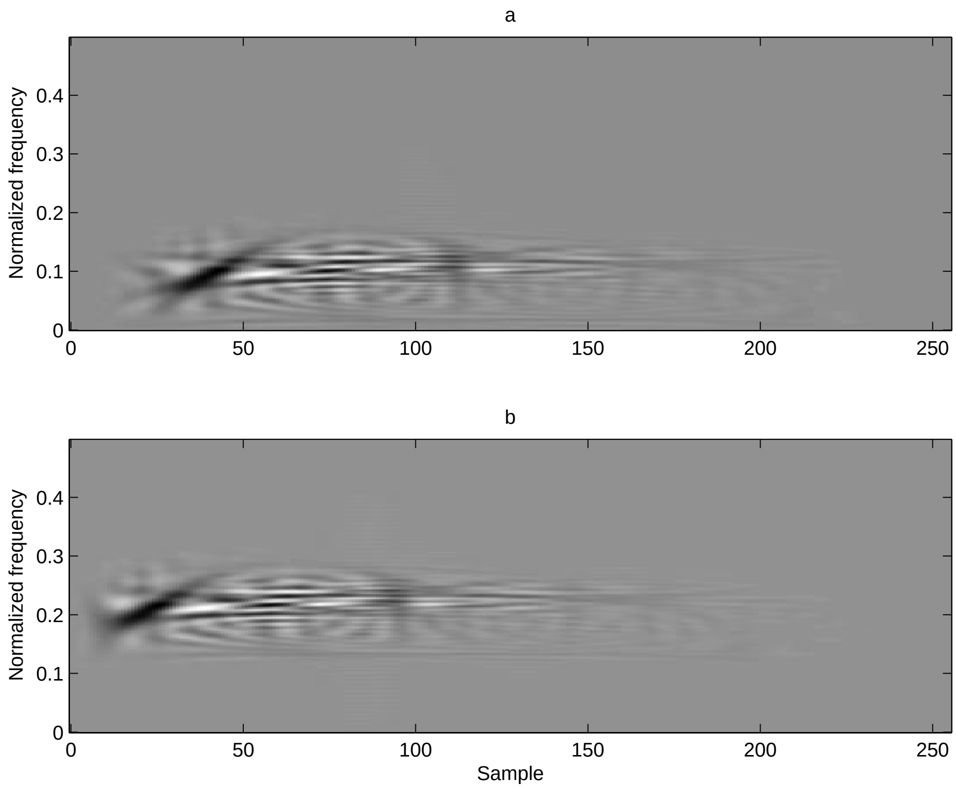


Figure 3. TFDs of a. original, b. frequency shifted click

6. APPLICATIONS

Several applications of the STIR/ZSS approach are presented in this section. There are two helicopter applications; the first application involves the detetion of faults in a helicopter transmission and the second application involves an attempt to detect and classify two radar backscatter signals, one from a two bladded rotating object and the other from a three bladed rotating object.

The next example is from a biomedical application and involves classification of temporomandibular joint (TMJ) clicks. The next example involves mammal sounds with emphasis on marine mammal sounds, but some comments on human speech. A word spotting problem represents a shift to images rather than TFDs and provides the final example of application of the transform methods.

6.1. Helicopter transmission analysis

This is the simplest example, since the focus is on application to a signal rather than a TFD. An interest in helicopter transmission diagnosis was expressed via an ONR initiative and a workshop held at Virginia Beach.[6] Participants were provided with data from several types of transmission faults. The goal was to detect and classifiy the fault. The STIR method was applied to these data with very good results. Various TFDs were applied and the binomial TFD performed best, in general. However, even better performance was obtained by simply using the spectra of the signals. The amplitude spectra were scale transformed and the magnitude of each transform was retained. In both cases the inner products of the spectra or scaled spectra were found to produce the scatter plot values shown. The first two principal vectors were used to produce the scatter plots. The results are shown in Fig. 7. Conventional methods using the principal vectors of the SVD to represent the results by using first two principal vectors produced the results shown in Fig. 7. However, when the ZSS method was used, the scaled spectra results were much better despite the better separation of ordinary spectra results shown in Fig. 7. Correct detection of faults rose dramatically.

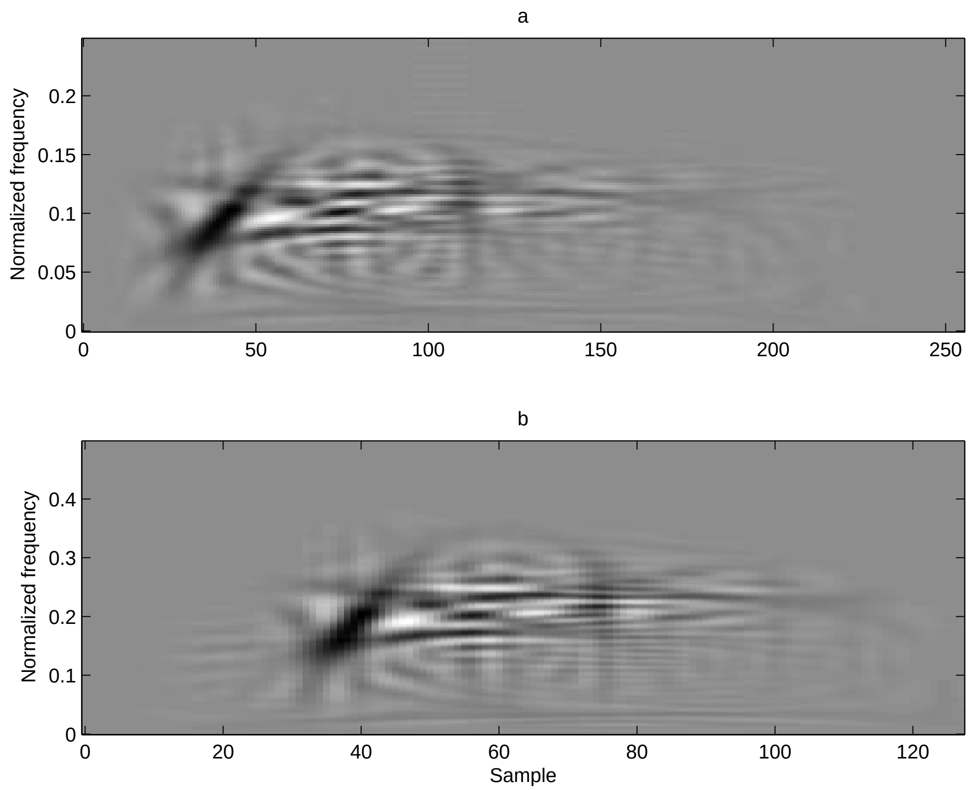


Figure 4. 'TFDs of a. original, b. scaled click

6.2. Helicopter radar backscatter analysis

This application involved an attempt to distinguish helicopter types by TFD analysis of simulated radar backscatter signatures.[7] Two bladed and three bladed simulations produced interesting Doppler signatures as shown in Fig 8. Signatures that varied in time-shift, frequency shift and scale were produced for each type of blade configuration. The final result included added noise as well. One hundred different examples were provided for each blade type. The binomial TFD performed best compared with Wigner, Born Jordan and spectrograms. The spectrogram performed well despite the fact that it is not scale covariant. The equal error rate (EER) for 3-bladed vs 2-bladed was about 5 % and the EER for detecting 2-bladed in zero dB noise was also about 5 % .

6.3. TMJ clicks

STIR methods were applied to temporomandibular joint (TMJ) click TFDs in an attempt to classify them into several distinct classes.[8] The ZSS method and nearest neighbor methods were compared. It was found the STIR methods performed very well in distinguishing certain types of clicks, but dropping the scale transform step (TIR method) performed best over all. This illustrates the fact that removing frequency shift effects is not always desirable if those frequency shifts act to distinguish classes. The methods were highly effective in classifying TMJ clicks and other associated sounds, leading to possible improved diagnosis.

6.4. Marine mammal sounds

Marine mammal sound analysis was an original motivation for this work. Dr. William Watkins of Woods Hole Oceanographic Institution has long been interested in improved analysis and classification of marine mammal

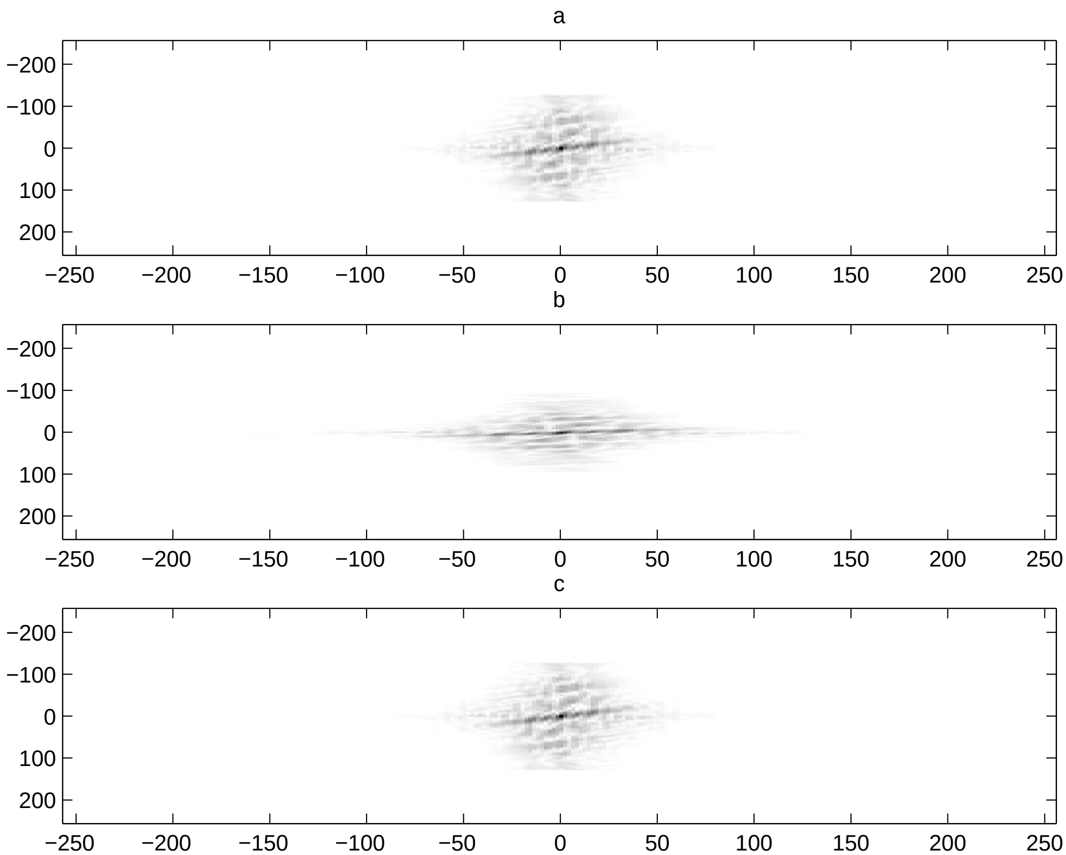


Figure 5. 2D FFTs of TFDS. a. original, b. scaled, c. frequency shifted

sounds. His trained ear is able to recognize different species of marine mammals and often, even different members of a family group. An ONR sponsored project provided the opportunity to work with Dr. Watkins in an attempt to improve the then used methods. In the context of this paper, success in recognizing individual sperm whales by means of their transient sounds has been accomplished.[9] The STIR method and an alternative method based on moment methods were studied. Others have found use in moment based methods for time-frequency classification.[10, 11]

6.5. Speaker verification

Success in classifying marine mammal sounds motivated attempts to use the methods for human speaker verification.[12] The results were generally successful and it was shown that these methods could perform almost as well as conventional speech methods. However, the TFD results are somewhat independent of conventional method results. This means that combining results could enhance speaker verification. Efforts in this arena are continuing under support from a NIST-ATP award to Quantum Signal LLC. A still unanswered question is the effect of removing pitch effects from speech, since pitch can be an important indicator of a given individual. If pitch effects are to be minimized in the STIR, then the full method is used. If not, then the frequency shift invariance step should be eliminated.

6.6. Word spotting

Some of the concepts reported here were refined in a word spotting project[13, 14] sponsored by NSA. The goal was to spot a given word in a faxed or otherwise corrupted document. Faxing documents creates unconventional

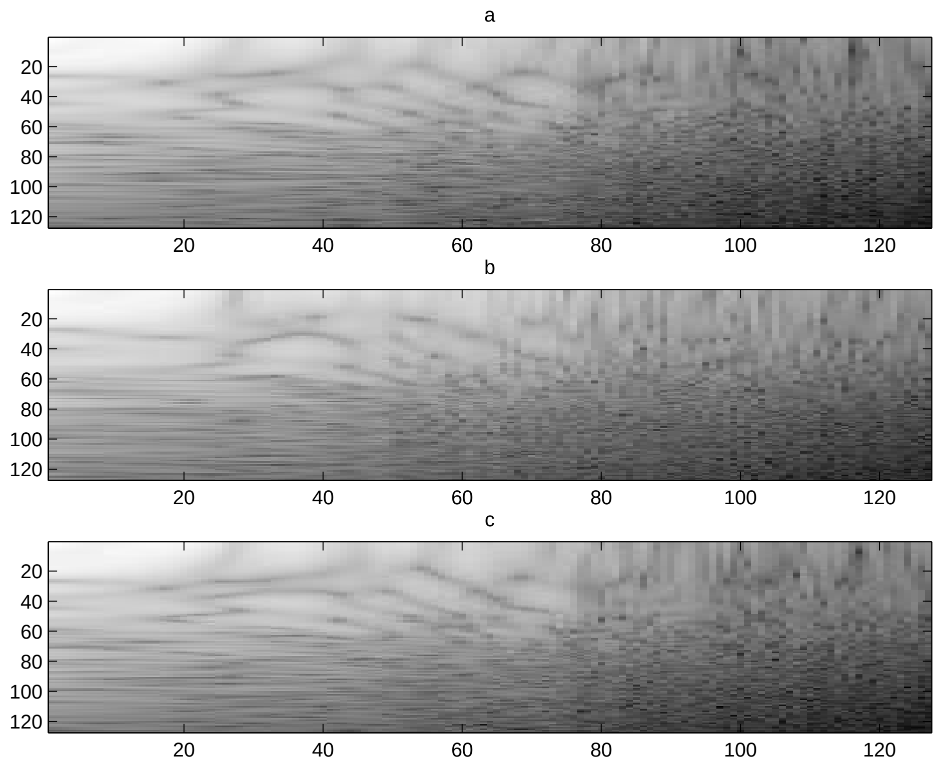


Figure 6. Final STIR result. a. original, b. scaled, c. frequency shifted

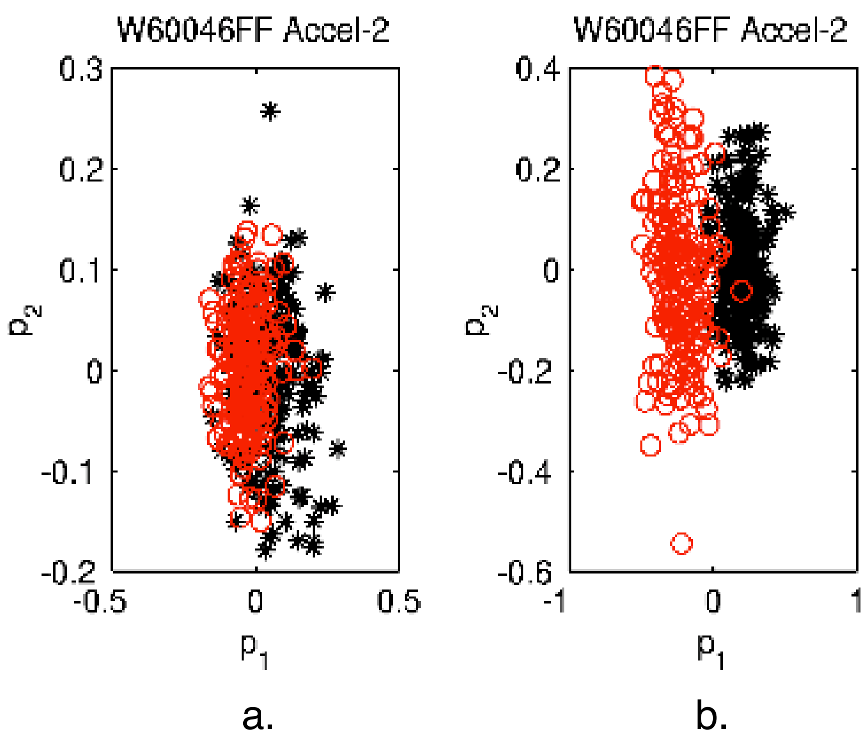
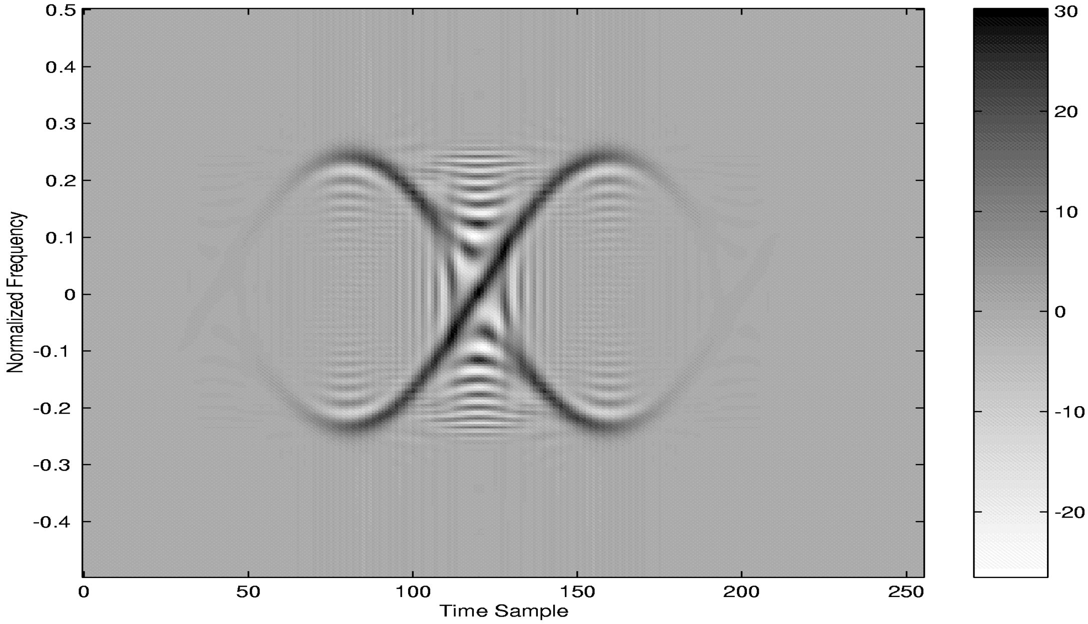
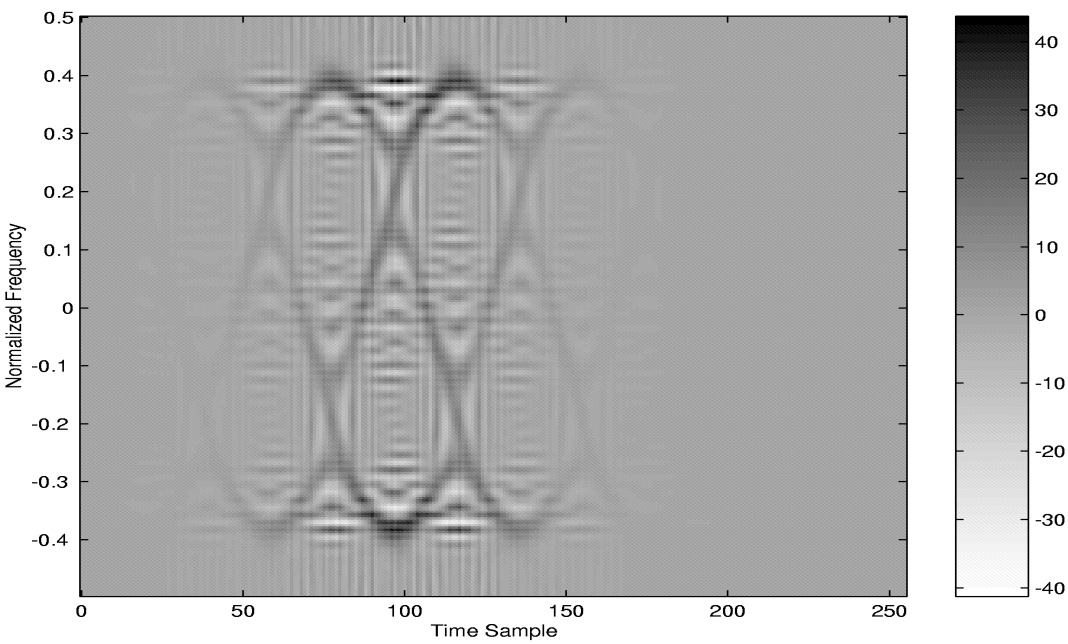


Figure 7. Scatter plots for a. scale transformed spectra, b. spectra with circles showing faults



a.



b.

Figure 8. Helicopter backscatter TFDS. a. Two bladed, b. Three bladed

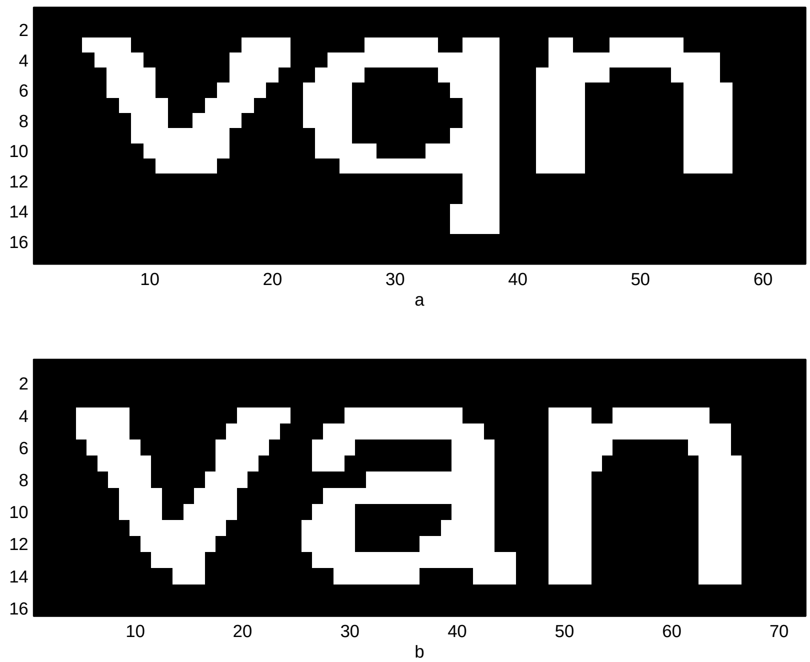


Figure 9. Two words. a. Incorrect word, b. correct word

noise evidenced by ragged edges in characters. The goal was to be able to spot a given work in a corrupted document for three fonts and for font sizes from 10 to 20 point. One does not need to produce a TFD in this application, though early work on this problem included spatial TFDs. The image of an isolated word can be taken through the STIR process, using the image rather than a TFD. Certain types of words were found to be the most troublesome. VAN was one such word. VAX was also troublesome and was sometimes confused with VAN. An example of fax corrupted VAN and another word with a "q" instead of an "a" is shown in Fig. 9. A number of examples of the desired words and others with one character replaced were produced and faxed. The recovered bitmaps were subjected to the STIR method. One can see the jagged edges caused by faxing. Sometimes, characters are altered such that they appear to be other characters. For instance, the middle letters in this example could be confused if enough parts were missing or altered. However, the STIR method produced about a 3 % EER, which is much better than is possible using conventional optical character recognition methods. The latter often produces 50 % errors under fax corruption.

6.7. Face recognition

It quickly became obvious that the word spotting method could be extended to images in general. Human face verification is of considerable current interest. Ongoing work sponsored by a NIST-ATP award is aimed at a significant improvement in this arena. Presently available methods often fail miserably under differences in lighting, pose and other variations. A great deal of progress beyond the original STIR method has been made.

7. CONCLUSIONS

It is clear that the STIR method and its ongoing enhancements and augmentations can solve a number of difficult problems. This is not the only promising method however. Moment based methods are also clearly of merit in the context of this work. In addition these methods can be combined with more conventional methods to improve their power. For example, neural net methods may be of considerable use in the final stage of classification. Clearly time-frequency analysis is being pushed beyond just "producing pretty pictures" by a number of groups at the present time. We are currently pursuing biometric applications of some of these concepts at Quantum Signal LLC.

Acknowledgements

This research was supported in part at the University of Michigan by the Office of Naval Research, ONR contract no. N00014-90-J-1654, and the National Science Foundation, NSF grant no. BCS-9110571 and NSA contract no. MDA904-95-C-2157. Work continues at Quantum Signal LLC under a NIST-ATP award for biometrics applications.

REFERENCES

1. L. Cohen, "Time–frequency distributions – a review," *Proceedings of the IEEE* **77**, pp. 941–981, 1989.
2. L. Cohen, "Introduction: a primer on time–frequency analysis," in *Time–Frequency Signal Analysis: Methods and Applications*, B. Boashash, ed., ch. 1, Longman Cheshire, Melbourne, Australia, 1991.
3. L. Cohen, "The scale representation," *IEEE Transactions on Signal Processing* **41**, 1993.
4. E. J. Zalubas and W. J. Williams, "Discrete scale transform for signal analysis," in *Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Processing*, **3**, pp. 1557–1561, 1995.
5. W. J. Williams, E. J. Zalubas, R. M. Nickel, and A. O. Hero, III, "Scale and translation invariant methods for enhanced time-frequency pattern recognition," *Multidimensional Signals and Systems* **9**, pp. 465–473, Oct. 1998.
6. W. J. Williams and E. J. Zalubas, "Helicopter transmission fault detection using time-frequency and scale techniques," *Journal of Mechanical Systems and Signal Processing* **14**(4), pp. 545–559, 2000.
7. W. J. Williams and E. J. Zalubas, "Invariant classifications of time frequency representations: applications to doppler radar target identification," in *Proceedings of the US/Australia Joint Workshop on Defence Applications of Signal Processing*, D. Cochran and B. Moran, eds., pp. 278–295, Elsevier, 2001.
8. D. Djurdjanovic, S. E. Widmalm, W. J. Williams, C. K. H. Koh, and K. P. Yang, "Computerized classification of temporomandibular joint sounds," *IEEE Transactions on Biomedical Engineering* **47**(8), pp. 977–984, 2000.
9. E. Zalubas, J. O'Neill, W. Williams, and A. O. Hero, III, "Shift and scale invariant detection," in *Proc. of the IEEE Int. Conf. on Acoust., Speech, and Signal Processing*, **5**, pp. 3637–3640, 1997.
10. B. Tacer and P. Loughlin, "Time-frequency based classification," in *Advanced Signal Processing Algorithms, Architectures and Implementations VI*, **2846**, pp. 186–192, Proceedings SPIE, (San Diego), Aug. 1996.
11. B. Tacer and P. Loughlin, "Nonstationary signal classification using the joint moments of time-frequency distributions," *Pattern Recognition* **31**(11), pp. 419–424, 1998.
12. R. M. Nickel and W. J. Williams, "On local time-frequency features of speech and their employment in speaker verification," *J. of the Franklin Institute* **337**, pp. 469–481, Apr. 2000.
13. W. J. Williams, E. J. Zalubas, and A. O. Hero, III, "Word spotting in bitmapped documents," in *Proc. 1997 Symposium on Document Image Understanding Technology*, pp. 214–227, May 1997. Annapolis, MD.
14. J. C. O'Neill, A. O. Hero, III, and W. J. Williams, "Word spotting via spatial point processes," in *IEEE International Conference on Image Processing*, pp. 217–220, 1996.