DISTANCE METRICS FOR DISCRETE TIME-FREQUENCY REPRESENTATIONS

James G. Droppo and Les E. Atlas

Department of Electrical Engineering
University of Washington
Seattle, WA
{jdroppo|atlas}@u.washington.edu

ABSTRACT

This paper presents a new method for addressing problems related to the sparsity of the discrete time-frequency representation. Although each real signal $x[n]$ has only N free parameters, its representation has N^2 elements, along $N(N+1)/2$ independent dimensions. This leads to problems in modeling, classification, and recognition tasks.

An overview of our discrete time-frequency representations is presented, together with a brief discussion of previous work in continuous time-frequency representations. These discrete time-frequency representations have all of the descriptive power of conventional discrete spectral features, together with a powerful framework that describes how the spectrum evolves over time. Unfortunately, using these features directly in their natural, high-dimensional space tends to be unworkable.

A geodesic distance measure is introduced, that leverages our knowledge of the set of valid time-frequency representations to reduce the apparent dimensionality of the problem. Applications of this geodesic distance are found in signal classification and nonlinear time-frequency interpolation.

1. INTRODUCTION

Standard techniques such as Gaussian mixture models, vector quantization, or k-nearest neighbor can break down when applied to discrete time-frequeny representations. There are three reasons for these difficulties. The first reason is data sparsity. For real signals with length N, the time-frequency representation consists of $N(N+1)/2$ linearly independent dimensions. To model this feature accurately using conventional methods, on the order of N^2 independent training examples would be needed, which is often impractical. The second reason is that not all configurations in this high-dimensional feature space are valid time-frequency representations. In particular, it can be shown that the set of all valid time-frequency representations is a continuous N dimensional surface embedded in the feature space. As a result, at every point in the feature space, a distribution has zero extent in at least $N(N-1)/2$ dimensions. The third problem with modeling with traditional techniques is that the mean feature may not lie near any valid representations. That is, for a given set of data, the mean of the data does not correspond with an actual signal. In other words, the mapping from the representation to the signal is over-determined.

As an example of the problems that may be encountered, consider the case of a set of three dimensional data restricted to lie on the surface of a cone. While it may take many mixtures of full covariance Gaussians to model the data accurately, there is a simple underlying structure that should be exploited. Section 2 presents an abbreviated derivation of the discrete time-frequency representations that are the focus of this paper, together with an overview of previous work on continuous time-frequency representations and a short discussion of how to read and interpret these representations. Section 3 describes what can be expected of the set of valid time-frequency representations. The dimensionality of the representations is explored, together with the gross shape of the set. Section 4 continues the discussion by introducing a geodesic distance metric that leverages the information contained in Section 3. Three methods are explored to compute and approximate this metric.

Applications in signal classification and non-linear interpolation are discussed in Sections 5 and 6. These Sections illustrate the usefulness of this geodesic metric for classification and interpolation tasks. The metric incorporates topological knowledge of the feature space, and improves performance for classification in addition to having some novel signal morphing, or non-linear signal interpolation, properties.

2. DISCRETE TIME-FREQUENCY REPRESENTATIONS

Our time-frequency representations are based on the work of Leon Cohen, who developed a systematic way of obtaining bilinear, joint bivariate densities of time and frequency using operator theory[1]. More recently [2], he presented a generalized approach for obtaining joint representations for arbitrary quantities (not just time and frequency) using the characteristic function operator methods of Moyal [3] and Ville [9].

Although Cohen's class of continuous quadratic TFRs are backed by rich theory, they have to undergo discrete

This work was supported by Microsoft Research and the Office of Naval Research.

sampling, resulting in discrete-time, discrete-frequency functions, before they can be implemented in any digital system and be of practical use. A considerable amount of work has gone into the study of sampling the continuous-time TFRs and ameliorating the ill effects introduced by the process, e.g. [5].

Our approach to formulating discrete-time, discrete-frequency Representations [4] avoids these sampling issues by providing a direct theoretical link between the discrete time sequence $x[n]$ and its discrete time-frequency representation $P[n,k]$. The derivation parallels Cohen's method for continuous time and frequency operators, but substitutes discrete operators in their place. As in the continuous case, each discrete time sequence is associated with not one, but a multitude of time-frequency distributions, each one uniquely specified by the root distribution and a kernel function.

A similar method of deriving discrete time-frequency representations, based on group theory, has been presented by Richman, *et al.* in [6]. One notable difference between the two is that whereas Richman's technique incorporates different calculations for even and odd length signals, our technique is invariant to signal length.

2.1. Derivation

We begin by assuming that we have a discrete, periodic sequence $x[n]$ with period N, whose discrete Fourier transform is $X[k]$. By developing discrete (matrix) operators corresponding to discrete time and frequency variables, we can generate a discrete representation in time and frequency, $P[n,k]$. This yields a technique that is distinct from sampling a distribution produced with the continuous time and frequency operators. There are problems inherent in sampling the continuous distribution that the discrete theory circumvents, by mapping directly from a discrete signal to a discrete time-frequency representation.

The first step is, of course, to formulate the discrete counterparts of the time and frequency operators. The time operator in the time domain, $\mathcal{L}$, and the frequency operator in the time domain, $\mathcal{K}$, obey the equations

$$\sum_{n=0}^{N-1} x^*[n]\mathcal{K}x[n] = \sum_{k=0}^{N-1} X^*[k]kX[k], \text{ and} \quad (1)$$

$$\mathcal{L}x[n] = nx[n]. \quad (2)$$

Since these operators are discrete and linear, they can be compactly represented by the matrices $\mathbf{L}$ and $\mathbf{K}$. Operator theory dictates that we can compute time-frequency representations according to the expectation

$$M[\eta,\tau] = \langle \exp(j2\pi\mathbf{L}\eta)\exp(j2\pi\mathbf{K}\tau)\rangle. \quad (3)$$

Simplifying Equation 3 above, and absorbing the constants into the operators for convenience, we get

$$\begin{aligned} & M[\eta,\tau] \\ = \; & \langle \exp(j\eta\mathbf{L})\exp(j\tau\mathbf{K})\rangle \\ = \; & \sum_{n=0}^{N-1} x^*[n]\exp(j\eta\mathbf{L})\exp(j\tau\mathbf{K})x[n+\tau] \\ = \; & \sum_{n=0}^{N-1} x^*[n]\exp(j\eta\mathbf{L})x[n+\tau]. \end{aligned} \quad (4)$$

A two dimensional discrete Fourier transform and further simplification yields the time-frequency representation,

$$P[n,k] = x^*[n]\exp(jnk)X[k], \quad (5)$$

which is the discrete version of the well-known Rihaczek TFR [7]. This result is analogous to the continuous-time result which can be derived in a similar fashion.

Just as in the continuous case, we can take advantage of the correspondence rule to represent permutations of the operators in Equation 3 as a multiplicative kernel function $\phi[\eta,\tau]$. In this way, the complete set of time-frequency representations within this class can be generated in the usual way. Members of this class of discrete time-frequency representations include the discrete Rihaczek TFR [7], the discrete Margeneau-Hill, and the spectrogram.

2.2. Interpretation

Time-frequency representations promise to capture both static spectral information and evolutionary spectral information in a single feature. Although representations in time and frequency usually make intuitive sense to even the untrained viewer, representations in η and τ, as in Equation 4, tend to be less intuitive.

The representation described by Equation 4 contains a lot of information about the signal $x[n]$. If $\eta = 0$, then $M[\eta,\tau]$ becomes the stationary autocorrelation of the signal $x[n]$, with τ taking its familiar role as the time lag variable.

$$M[0,\tau] = \sum_{n=0}^{N-1} x^*[n-\tau]x[n] \quad (6)$$

As a result, values along the $\eta = 0$ axis are always interpreted as the autocorrelation of the signal $x[n]$. Conversely, any stationary signal will concentrate its energy within this region.

If $\tau = 0$, then $M[\eta,\tau]$ becomes the spectrum of the instantaneous energy of the signal $x[n]$. This leads us to refer to η as the modulation frequency variable, because it relates to how quickly the envelope of the signal is changing.

$$M[\eta,0] = \sum_{n=0}^{N-1} |x[n]|^2 \exp\left(\frac{j2\pi\eta n}{N}\right) \quad (7)$$

As a result, values along the $\tau = 0$ axis show how the modulation "envelope" of the signal $x[n]$ is changing with time. A signal with=20 little correlation between its samples will concentrate its energy in this region.

Any point $M[\eta,\tau]$ refers to how quickly a specific correlation coefficient is being modulated. It can be interpreted as the output of a non-stationary sinusoidal filter-bank, where distance from the origin is analogous to bandwidth, and the angle from the η axis represents the linear chirp rate of the filter-bank.

To see this, consider a kernel function

$$\phi[\eta,\tau] = \delta[\eta-a]\delta[\tau-b], \text{ where} \quad (8)$$

$$\delta[n] = \begin{cases} 1 & n=0 \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

In the $(n \times k)$ plane, this kernel becomes a complex exponential,=20 with parameters controlled by a and b.

$$\Phi[n,k] = \exp\left(-j2\pi(an+bk)\right) \tag{10}$$

In summary, a stationary process has no extent in η, a spectrally flat process has no extent in τ, and more interesting structures appear when $\eta \neq 0$ and $\tau \neq 0$. Time modulations increase a signal's extent in η, and colored spectra are responsible for extent in τ.

2.3. Unleashing Discrete Time-Frequency Representation's Power

The traditional disadvantage to using these representations as features is the deluge of information. Whereas stationary spectral features produce vectors with at most a few dozen dimensions, a real length N signal will always produce a representation with $N(N+1)/2$ linearly independent dimensions. For a 250 ms signal sampled at 16 kHz, this amounts to no less than eight million dimensions.

There have been several attempts in the literature to reduce this information into salient features, including taking autocorrelation coefficients of the representation in time or frequency to achieve a time-shift or frequency-shift invariant representation of the data [10], and finding two dimensional moments of the time-frequency representation [8].

Given the apparently immense dimensionality of the feature space, one might assume that extracting useful information in an automated and efficient way would be impossible. This is not the case. The next sections present a new way of looking at these representations.

3. THE SET OF VALID TIME-FREQUENCY REPRESENTATIONS

Consider the discrete-time signal $x[n]$, with length N. The discrete time-frequency representation $M[\eta,\tau]$ associated with $x[n]$ consists of N^2 complex numbers, and is contained in the set of length N^2 complex vectors.

$$M[\eta,\tau] = \sum_{n=0}^{N-1} x^*[n]x[n+\tau]\exp\left(2\pi\eta n/N\right).$$

This, however, is somewhat misleading as to the true size of the feature space. A discrete Fourier transform from η into n yields,

$$A[n,\tau] = x^*[n]x[n+\tau]. \tag{11}$$

Clearly, there are $N(N-1)/2$ unique products when $\tau \neq 0$, and N magnitudes when $\tau = 0$. For complex signals, this yields a feature space with N^2 linearly independent dimensions, and for real signals, the feature space has $N(N+1)/2$ linearly independent dimensions.

Furthermore, the mapping between the signal $x[n]$ and its representation $M[n,k]$ is continuous, and therefore the set of representations that correspond to actual signals forms a continuous subspace.

For real signals with $N = 2$, the set of valid time-frequency representations is a 90 degree cone embedded in a four-dimensional space, with extent only in three dimensions. This agrees with the predicted value, since the expected dimensionality of the cone is $2(2+1)/2 = 3$. This case is examined in detail in Section 4.1.

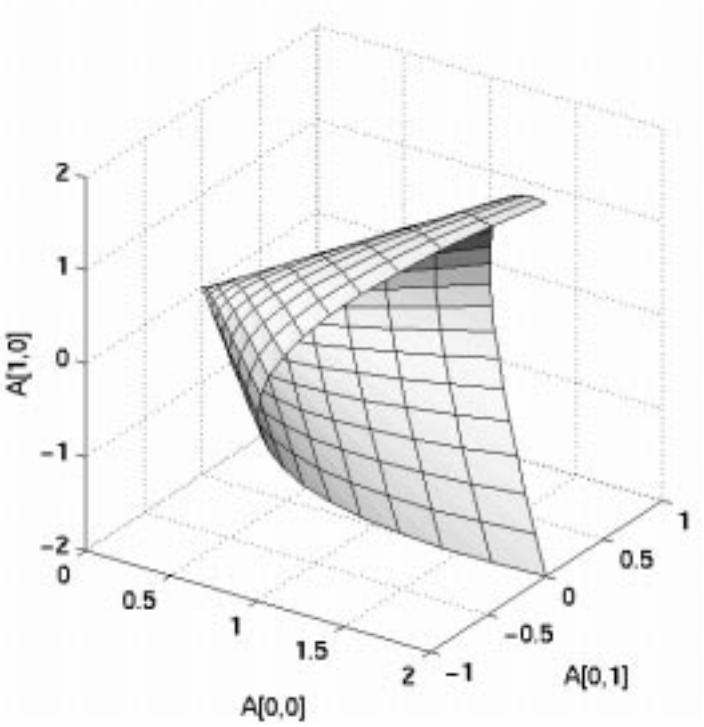

Figure 1: Conical structure of valid TFR

For larger values of N, the surface can not be named, but it can be functionally described. The set of valid discrete time-frequency representations always lies on a conical structure, regardless of N. The axis of the cone corresponds to the energy of the signal $x[n]$, which is stored at the origin of the (η,τ) plane, $M[0,0]$. For a fixed energy, the valid representations lie on a hyper-sphere with a radius proportional to $M[0,0]$.

$$\begin{aligned}\sum_{m=0}^{L-1}\sum_{n=0}^{L-1} = 20\left|A[m,n]\right|^2 &= \sum_{m=0}^{L-1}\sum_{n=0}^{L-1} = 20\left|x[m]x^*[n]\right|^2 \\ &= \sum_{m=0}^{L-1}\left|x[m]\right|^2\sum_{n=0}^{L-1}\left|x[n]\right|^2 \\ &= \left|M[0,0]\right|^2 \end{aligned} \tag{12}$$

Although all valid representations lie on this conical structure, not all points on this structure are valid representations. That is, meeting Equation 12 is a necessary, but not sufficient test for valid time-frequency representations.

4. A GEODESIC DISTANCE MEASURE

Measuring the distance between New York and San Francisco by drawing a straight line through three dimensional

space would not seem to be an acceptable measure. Similarly, measuring the distance along a line connecting two valid representations, but lying entirely outside of the surface of valid representations, may not be the best solution. This is exactly what is going on with a Euclidean distance metric in the time-frequency representation space.

The structure of the surface of valid time frequency representations, along with the calculus of variations, allows us to find the shortest line connecting two representations, that remains entirely within the set of representations of true signals. This line is a geodesic, and its length is the geodesic distance between the endpoints.

This line can be used to morph one signal into another, along the shortest path in quadratic time-frequency space. That is, given two signals, it specifies a continuous path along which one signal slowly changes into the other. This idea is explored in Section 6. The line can also be used to find the "true" distance between distributions, which would not necessarily be a straight line in the feature space.

Without explicitly solving the geodesic problem, approximate solutions can be found. Two solutions for the geodesic distance are presented, an algorithmic approximation and a numerical approximation.

4.1. Exact Solution for Length Two Signals

This section deals with the case of real signals with $N = 2$. As mentioned previously, the set of valid representations for this case is a three dimensional cone embedded in a four dimensional space. An exact solution for the geodesic between two given endpoints can be found using the calculus of variations.

For convenience, let a and b represent the two samples of the signal $x[n]$. In the (η, τ) plane, this signal's representation only has three non-zero values.

$$\begin{array}{cc} M[0,1] = 2ab & M[1,1] = 0 \\ M[0,0] = a^2 + b^2 & M[1,0] = a^2 - b^2 \end{array}$$

A simple change of variables reveals the traditional formula for a cone,

$$\begin{aligned} x &= M[0,1] = u \sin v = 20 \\ y &= M[1,0] = u \cos v \\ z &= M[0,0] = u, \text{ where } u \geq 0 \end{aligned}$$

To determine the distance between two representations, we measure the shortest line along the surface that joins the two points. This is exactly the type of problem that the calculus of variations is meant to solve.

The problem is posed as one of minimizing the line integral between the two points. The calculus of variations tells us that the geodesic function $v(u)$ is given by

$$v = c_1 \int \frac{\sqrt{P}}{\sqrt{R(R - c_1^2)}} du,$$

where

$$P = \left(\frac{\delta}{\delta u} x\right)^2 + \left(\frac{\delta}{\delta u} y\right)^2 + \left(\frac{\delta}{\delta u} z\right)^2$$

$$Q = \left(\frac{\delta}{\delta u} x\right)\left(\frac{\delta}{\delta v} x\right) + \left(\frac{\delta}{\delta u} y\right)\left(\frac{\delta}{\delta v} y\right) + \left(\frac{\delta}{\delta u} z\right)\left(\frac{\delta}{\delta v} z\right)$$

$$R = \left(\frac{\delta}{\delta v} x\right)^2 + \left(\frac{\delta}{\delta v} y\right)^2 + \left(\frac{\delta}{\delta v} z\right)^2$$

and

$$\begin{array}{cc} \frac{\delta}{\delta u} x(u,v) = \sin(v) & \frac{\delta}{\delta v} x(u,v) = u\cos(v) \\ \frac{\delta}{\delta u} y(u,v) = \cos(v) & \frac{\delta}{\delta v} y(u,v) = -u\sin(v) \\ \frac{\delta}{\delta u} z(u,v) = 1 & \frac{\delta}{\delta v} z(u,v) = 0 \end{array}$$

so

$$P = 2 \quad Q = 0 \quad R = u^2$$

Combining these equations, the result is

$$\begin{aligned} v &= c_1 \int \frac{\sqrt{2}}{\sqrt{(u^2)(u^2 - c_1^2)}} du \\ &= c_1 \sqrt{2} \int \frac{1}{u\sqrt{u^2 - c_1^2}} du \\ v &= \sqrt{2} \cos^{-1} \left| \frac{c_1}{u} \right| + c_2 \end{aligned}$$

The constants c_1 and c_2 are dependent on the desired endpoints of the geodesic, (u_1, v_1) and (u_2, v_2).

$$\begin{aligned} c_1 &= \pm u \cos(\frac{v - c_2}{\sqrt{2}}) \\ c_2 &= v - \sqrt{2} \cos^{-1} \left| \frac{c_1}{u} \right| \\ c_2 &= -\sqrt{2} \tan^{-1} \left(\frac{u_1 \cos \frac{v_1}{\sqrt{2}} - u_2 \cos \frac{v_2}{\sqrt{2}}}{u_1 \sin \frac{v_1}{\sqrt{2}} - u_2 \sin \frac{v_2}{\sqrt{2}}} \right) \end{aligned}$$

4.2. Algorithmic Approximation

To find the shortest distance between two points, lying entirely along valid time-frequency representations, an initial solution is posed, and then iteratively re-estimated. This algorithm produces both an estimate of the geodesic distance, and evenly spaced points along the geodesic.

The initial estimate is to form a set of points equally spaced in the signal space, with the proper endpoints. If one endpoint is the signal $a[n]$, and the second endpoint is $b[n]$, the initial set of points in the signal space would be,

$$x_i[n] = a[n] + (b[n] - a[n]) \frac{i}{M - 1}, \text{ for } 0 \leq i < M.$$

The re-estimation assumes that links between neighboring signals are elastic, and tries to pull each signal towards its neighbors in the time-frequency representation space.

First, a direction is computed that will move a signal closer to its neighbors in the time-frequency representation space. If the representation for $y[n]$ moves in this direction, the distance to the representations $x[n]$ and $z[n]$ tends to decrease.

$$v[n, \tau] = \frac{1}{2}(x[n]x[\tau] + z[n]z[\tau]) - y[n]y[\tau]$$

This vector is then mapped to an equivalent direction in signal space. The derivatives of the representation with respect to the original signal samples are

$$\frac{d}{dy[a]} y[n]y[\tau] = \delta[n - a]y[\tau] + \delta[\tau - a]y[n].$$

Now, this is a scalar-valued function of n, τ, and a. Unwrap it into a matrix, whose columns are a, and whose rows consist of all combinations of n and τ. Call this matrix V. Its columns represent the directions the representation will move with changes in the signal values.

Now, all that remains is mapping the desired vector v, which exists in the representation space, onto the available directions contained in V. This is a simple matrix operation, and consists of the minimum square solution to the linear equation,

$$\begin{aligned} Vd &= v, \\ d &= (V^H V)^{-1} V^H v. \end{aligned}$$

If the signal y is moved in the direction d, the distances between the representations for y and z, and y and x, is reduced. If the representations of x, y, and z are sufficiently close, the linear distance and the geodesic distances should be approximately equal.

4.3. Numerical Approximation

The numerical approximation to the geodesic distance consists of first defining a path in the representation space, and then precisely computing its length.

According to Equation 11, the surface of valid time-frequency representations consists of N^2 dimensions y_{ij}, where $0 \leq \{i,j\} < N$. This surface is parameterized in terms of the independent variables x_i, where $0 \leq i < N$, which are the discrete signal samples.

$$y_{ij} = x_i x_j$$

The first step is to assume the signal values x_i are functions of a single independent variable, t. As t varies from 0 to 1, a path is traced through both the signal space and the representation space.

Let $\hat{u}_{ij}$ be a unit vector in the N^2 time-frequency representation space such that

$$\langle \hat{u}_{ab}, \hat{u}_{cd} \rangle = \delta[a-c]\delta[b-d].$$

The derivative of the path with respect to the independent variable is given by

$$\begin{aligned} \frac{d}{dt}\hat{s} &= \sum_{i,j} \frac{d}{dt} y_{ij} \hat{u}_{ij} \\ &= \sum_{i,j} \frac{d}{dt} x_i(t) x_j(t) \hat{u}_{ij} \\ &= \sum_{i,j} \left(x_j(t)\dot{x}_i(t) + x_i(t)\dot{x}_j(t) = 20 \right) \hat{u}_{ij} \end{aligned}$$

To find the distance along the path, integrate the function $f(t)$, where

$$\begin{aligned} f(t) &= \left| \frac{d\hat{s}}{dt} \right| \\ &= \left(\sum_{i,j} \left(x_j(t)\dot{x}_i(t) + x_i(t)\dot{x}_j(t) = 20 \right)^2 \right)^{\frac{1}{2}} \end{aligned}$$

In many dimensions, it is difficult to solve this equation exactly for the functions $x_i(t)$, $0 \leq i < N$, but an approximate solution can be found by assuming a linear form for the functions. The slope and offset for each function are entirely specified by the initial value, $x_i(0)$, and the final value, $x_i(1)$.

$$\begin{aligned} x_i(t) &= x_i(0) + (x_i(1) - x_i(0))\, t \\ \dot{x}_i(t) &= x_i(1) - x_i(0) \end{aligned}$$

Once the endpoints for the calculation are known, the function $f(t)$ takes the form of a square root of a quadratic function. This can be quickly estimated with great precision by calculating the integral

$$d = \int_0^1 f(t)\, dt.$$

This approximation is equivalent to the starting point of the iterative solution presented previously. Although this approximation is rather gross, in practice the differences between the two solutions do not amount to much. That is, if you draw a straight line between two signals in the signal space, the length of the corresponding line in the time-frequency representation space is close to optimal.

5. CLASSIFICATION IMPROVEMENT

Figure 2 shows the performance of three distance metrics for classification in additive white Gaussian noise. The "L2-Signal" metric is equivalent to the matched filter solution, and serves as an upper bound on performance. The "Geodesic Approximation" corresponds to a classifier using the numerical approximation to the geodesic distance. The "L2-Representation" classifier uses the distance from the mean in representation space.

The geodesic approximation does much better than computing distances directly in the representation space. It comes close to being the optimal detector, although it is just an approximation to the true geodesic distance.

6. NON-LINEAR INTERPOLATION, MORPHING

The points along a geodesic correspond to signals that are *between* the endpoints and *close* to neighboring points in the geodesic. By defining *between* and *close* in the discrete time-frequency representation space, it is reasonable to assume that as one travels along the geodesic, the signal's time and frequency structure are both changing.

If the endpoints are chosen to be similar signals that differ by a time shift, the algorithmic approximation presented in Section 4.2 converges to a sequence of similar signals, with an increasing time shift.

If the endpoints are chosen to be impulse response functions for digital filters with two different resonances, the algorithmic approximation converges to a sequence of signals, whose resonance is shifting from being close to the starting signal, to being close to the ending signal.

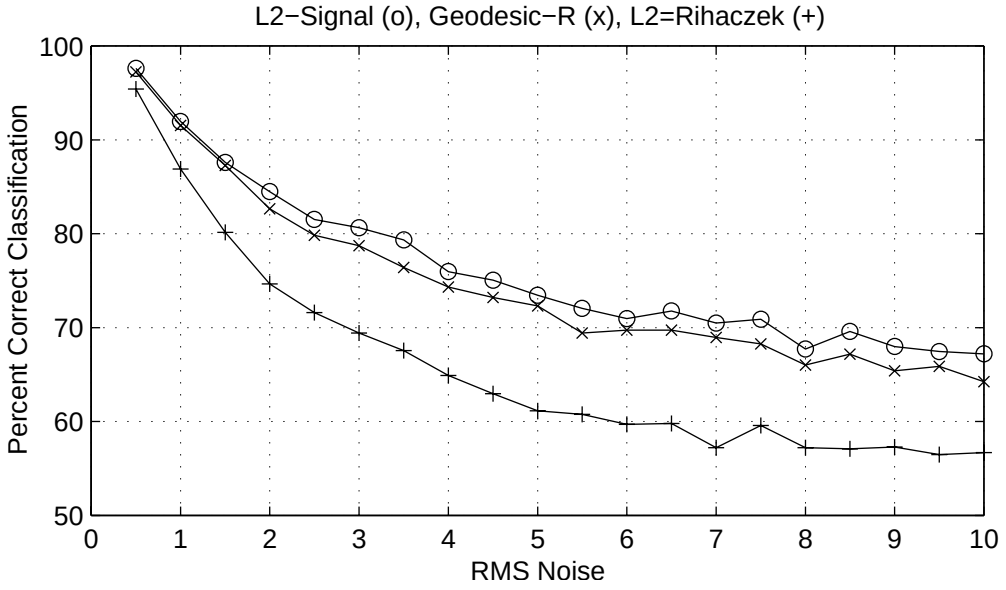

Figure 2: Classification in the presence of additive noise

7. CONCLUSION AND FUTURE WORK

This paper presented a new way of computing distances and paths between discrete time-frequency representations. A geodesic is computed, which incorporates topological knowledge of the feature space to reduce the apparent dimensionality of the problem.

Using the geodesic distance, a classifier was constructed that outperforms classifiers that do not incorporate any knowledge of the feature space. Furthermore, if the geodesic in the time-frequency space is used as a non-linear interpolation algorithm, reasonable results are obtained.

The kernel function $\phi[\eta, \tau]$ is almost completely ignored in this work. It could be re-introduced before beginning the geodesic calculation. Since it is a multiplicative mask on $M[\eta, \tau]$, the net effect would be a cost function on the path moving along directions with high ϕ values. One could use the kernel to impose no cost for changing the modulation of the signal, or vary the cost along τ to penalize finer modifications to the spectrum more or less.

REFERENCES

[1] Leon Cohen. Generalized phase-space distribution functions. *J. Math. Phys.*, 7:781–786, 1966.

[2] Leon Cohen. *Time-Frequency Analysis*. Prentice Hall Signal Processing Series, New Jersey, 1995.

[3] J. E. Moyal. Quantum mechanics as a statistical theory. In *Proc. Camb. Phil. Soc.*, volume 45, pages 99–124, 1949.

[4] Siva Bala Narayanan, Jack McLaughlin, Les Atlas, and James Droppo. An operator theory approach to discrete time-frequency distributions. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pages 521–524, 1997.

[5] A. H. Nutall. Alias-free Wigner distribution functions and complex ambiguity functions for discrete-time samples. Technical Report 8533, Naval Underwater Systems Center, April 14 1989.

[6] M. Richman, T. Parks, and R. Shenoy. Features of a discrete Wigner distribution. In *1996 IEEE Digital Signal Processing Workshop Proceedings*, pages 427–430, 1996.

[7] A. W. Rihaczek. Signal energy distribution in time and frequency. *IEEE Transactions on Information Theory*, 14:369–374, 1968.

[8] B. Tacer and P. Loughlin. Time-frequency based classification. In *Proceedings of the SPIE–The International Society for Optical Engineering*, volume 2846, pages 186–92, Denver, CO, USA, August 1996.

[9] J. Ville. Theorie et applications de la notion de signal analytique. *Cables et Transmission*, 2 A:61–74, 1948.

[10] E. J. Zalubas, J. C. Oneill, W. J. Williams, and A. O Hero III. Shift and scale invariant detection. volume 5, pages 3637–40.