An Introduction to the Multiresolution Fourier Transform and its Applications

R. Wilson, A.D. Calway, E.R.S. Pearson and A.R. Davies.

Abstract

The Multiresolution Fourier Transform (MFT) is a new tool for signal analysis, which is designed with express intent of providing a *local* signal representation adapted to the problem of segmenting signals into a set of meaningful *primitive* features. It is shown that this requires a signal representation that is richer than conventional Short Time Fourier Transforms (STFT) or indeed Wavelet Transforms (WT), but shares with them the property of giving a *phase-space* description of the signal. These considerations lead to the new transform, which can be seen as a marriage of the STFT and WT which overcomes the limitations of either as a signal representation for segmentation. Some of the elementary properties of the MFT are then discussed and its implementation for discrete signals is considered. The report is concluded with a summary of results obtained with the MFT on audio and image signal segmentation.

Index Terms: Fourier Analysis, Multiresolution, Wavelet Transform, Uncertainty, Markov Models, Image and Audio Segmentation.

1 Introduction

In recent years, there has been growing interest in *local* signal representations, whose *basis* or *frame* vectors [1] are effectively confined to some compact region of support. Although they have appeared in the literature in a number of guises, many of these can be placed into one of two classes : Short-time Fourier Transforms (STFT) [10, 12, 11] or Wavelet Transforms (WT) [4, 5, 3, 7, 6, 1]. Closely related to the former is the Gabor representation [8, 46] , while a variety of pyramid and scale-space methods are more directly linked to the WT [2, 20, 47]. Both classes are *phase-space* representations : they provide a co-ordinate system in which a 'property' co-ordinate, respectively frequency (STFT) or scale (WT) is *conjoined* to position. Correspondingly, in the continuous case, the set is closed under the group of translations and either translations in frequency (STFT) or dilations (WT) [1].

One important application for such representations is what may be called 'perceptual signal analysis' : the identification from a complex 'real world' signal of a relatively small number of perceptually significant signal features, such as the notes in a piece of music [39, 37, 38, 17] or the boundary contours of objects in a scene [21, 15, 16]. Because such features are of finite extent in time or space and because there may be several in any given compact region, conventional Fourier analysis is of limited use in such problems : it is an idealisation, not a practical tool [9, 12]. Indeed, these factors tightly constrain the set $\mathcal{G}$ of vectors used to represent the features : if Ω is any compact region, there should exist a subset $\mathcal{G}(\Omega)$ of vectors whose effective support is Ω, such that any signal whose support is confined to Ω can be represented as a linear combination of vectors in $\mathcal{G}(\Omega)$. This is the only way to guarantee that an arbitrary signal feature can be extracted without interference from neighbouring features, regardless of their size, position or type. The STFT and WT represent alternative approaches to the localisation problem, which nonetheless share a significant feature - in both cases the group of translations assumes a primary role. In the context of segmentation, this is indeed desirable - it is almost an unwritten law that changes of position do not affect the classification of a signal feature; a musical note remains the same note if it is shifted in time; similarly, identification of visible objects should not be affected by changes in their positions. The closure of the set of frame vectors of both STFT and WT under translations is just the mathematical expression of this necessary property. Unfortunately, although they provide partial solutions to the localisation problem, neither the STFT nor the WT represents a completely general approach : for the STFT to be effective, the neighbourhood size must be known *a priori*, while the in the WT it is not the size of the region, but its content which is constrained by the locality requirement (the set of WT vectors with support confined to a given region is not in general a basis for the set of signals with support is confined to that region). In both cases, therefore, there is a constraint on the signal features which can be easily segmented - a restriction on their domains of application. This restriction is no accident - it is an inevitable consequence of the fact that both representations are *frames* (cf section 2.1).

The work reported here is an attempt to grapple with this problem, based on a new linear transform - the Multiresolution Fourier Transform (MFT) [14, 15]. The MFT is a superset of the STFT and WT, which has the ability to represent arbitrary signals of compact support in an interference-free manner. In the next section, a novel approach to the problem of segmentation is presented, in which scale is used to overcome the limitations of a fixed window. This allows the derivation of significant constraints on the associated signal representations, which are then used to identify the MFT as the unique linear transform suited to the segmentation of arbitrary signals of compact support. The MFT is then defined for various signal domains and some of its elementary properties are given. A brief account is then presented of its application to the segmentation of audio signals and images [17, 15, 19]. These applications make use of hierarchical Markov models operating in the MFT domain. The paper is concluded with a summary of the main features of the work and some suggestions for its extension.

2 Towards the MFT

2.1 Scale in segmentation

It may be helpful to begin by explaining how systematic variation of scale can simplify the segmentation problem : deciding on the basis of the signal $x(\vec{\chi}), \vec{\chi} \in \mathcal{N}(\vec{\xi})$, in a neighbourhood of the point $\vec{\xi}$ that an 'object' X is present at that point and estimating its parameters $\boldsymbol{\phi}_X(\vec{\xi})$. This suggests the use of a phase-space description : the X's must be defined in terms of properties which can be estimated from the signal in the neighbourhood $\mathcal{N}(\vec{\xi})$. In other words, the point $\vec{\xi}$ should be replaced by the phase-space co-ordinate $(\vec{\xi}, \vec{\zeta})$, the latter being a point in some property domain, representing those aspects of $x(\vec{\xi})$ in the neighbourhood relevant to the segmentation. Because uncertainties surround the inference, it is normally cast in statistical terms [34], based on the posterior probabilities of the parameters and the hypothesis $H_X(\vec{\xi})$ that there is an X at $\vec{\xi}$

$$P(\boldsymbol{\phi}_X(\vec{\xi}), H_X(\vec{\xi}) | x(\vec{\chi}), \vec{\chi} \in \mathcal{N}(\vec{\xi})) =$$
$$\text{Prob}\{X \text{ at } \vec{\xi} \text{ with parameters } \boldsymbol{\phi}_X(\vec{\xi}), \text{ given } x(\vec{\chi}),\ \vec{\chi} \in \mathcal{N}(\vec{\xi})\} \qquad (1)$$

The problem is one of maximising these conditional probabilities with respect to the choice of hypothesis and parameters. It often happens that the prior distributions are unknown, in which case an ML approach may be used - the posterior probabilities of (1) are replaced by the conditionals $P(x(\vec{\chi}), \vec{\chi} \in \mathcal{N}(\vec{\xi}) | \boldsymbol{\phi}(\vec{\xi}), H_X(\vec{\xi}))$. An obvious defect of this formulation is the presence of the unspecified neighbourhood $\mathcal{N}(\vec{\xi})$: if it is too small, then lack of relevant data will adversely affect the decision; if it is too big, then there may be several objects $X_1, X_2..$, in the window, so that the set of plausible hypotheses is apt to undergo a combinatorial explosion - it has to include all likely combinations of X's. An alternative to maintaining a fixed window and varying the set of hypotheses is to fix the set of hypotheses and vary the window scale : to use a multiresolution representation (eg [20, 21, 30, 35]). The inferences made at different scales may then cross-checked for *scale consistency* : what can be found at a scale σ should not vary rapidly as σ is varied, if it is to be believed. In other words, assuming that the X's are a finite distance apart in phase-space, there will be a range of scales over which there will only be one X in the neighbourhood of a given point $(\vec{\xi}, \vec{\zeta})$ in the phase-space. This can be expressed in the ML formulation using

$$\begin{aligned} P(\hat{x}(\vec{\xi}, \vec{\zeta}, \sigma_1), \sigma_1 \geq \sigma | \boldsymbol{\phi}_X, H_X) = P(\hat{x}(\vec{\xi}, \vec{\zeta}, \sigma_1), \sigma_1 > \sigma | \hat{x}(\vec{\xi}, \vec{\zeta}, \sigma), \boldsymbol{\phi}_X, H_X) \\ \times P(\hat{x}(\vec{\xi}, \vec{\zeta}, \sigma) | \boldsymbol{\phi}_X, H_X) \end{aligned} \qquad (2)$$

via Bayes's law. The use of $\hat{x}$ in (2) indicates that the phase-space transformed data are being used. The use of scales $\sigma_1 \geq \sigma$ is justified by the observation that, for any realisation of the signal, there must be some largest scale (ie smallest σ) for which there is at most one X in the window around $\vec{\xi}$. This scale need not be known *a priori* : the significant point is that the phase-space representation at larger scales will not contribute useful data to the inference. Were this not the case, the phase-space could be replaced by the representation $\hat{x}(0, \vec{\zeta}, 0)$. The conditional probabilities of (2) suggest the sequential procedure :

1. Perform the inference separately at each scale σ, obtaining an initial decision about H_X and estimates of parameters $\boldsymbol{\phi}_X$, using $P(\hat{x}(\vec{\xi}, \vec{\zeta}, \sigma) | \boldsymbol{\phi}_X, H_X)$ The parameter estimates derived from maximising this probability will be called $\bar{\boldsymbol{\phi}}_X(\sigma)$.

2. Test for scale consistency, based on the conditional probabilities of the data at scales $\sigma_1 > \sigma$, given the data at σ, the hypothesis and parameters, using the fact that the parameter estimates $\bar{\boldsymbol{\phi}}_X(\sigma)$ are now given, so that

$$P(\hat{x}(\vec{\xi}, \vec{\zeta}, \sigma_1) | \boldsymbol{\phi}_X, H_X) = P(\hat{x}(\vec{\xi}, \vec{\zeta}, \sigma_1) | \bar{\boldsymbol{\phi}}_X(\sigma_1), H_X) P(\bar{\boldsymbol{\phi}}_X(\sigma_1) | \boldsymbol{\phi}_X, H_X) \qquad (3)$$

The first term on the *r.h.s.* of (3) is the conditional used in the initial inference at scale σ_1, evaluated at the best parameter estimate, while the second is the scale consistency. It provides both an improved parameter estimate and a better assessment of the likelihood that there is an X at $\vec{\xi}$. This step can be repeated over scales

$\sigma_m > \sigma_{m-1} > .... > \sigma$, until an acceptable level of confidence is reached or a maximum scale is reached, at which the data can be shown to carry no useful information. The obvious benefit of this approach, over a fixed window method, is that it allows users to maintain a fixed set of hypotheses without penalising their ability to make reliable inferences about them. Such procedures demand, however, that the set of vectors used to represent the signal can represent any signal whose support is confined to a particular neighbourhood without interference from signals supported in any mutually disjoint neighbourhood. This leads to the following definitions, which apply to $1-d$ signals belonging to the Hilbert space $L^2(R)$.

Definition. A set $\mathcal{G} = \{g_n, n \in Z_+\}$ of vectors $g_n \in L^2(R)$ is said to be *locally complete* in $L^2(R)$ if each g_n is of compact support and if, for each finite interval $[\xi_1, \xi_2)$, there exists a subset of $\mathcal{G}$, $\mathcal{G}(\xi_1, \xi_2) = \{g_n, g_n(\xi) = 0, \xi < \xi_1, \xi \geq \xi_2\}$, of vectors whose support is contained in that interval, which is complete in $L^2(\xi_1, \xi_2)$.

An immediate consequence of the definition is that $\mathcal{G}$ is complete in $L^2(R)$. Although necessary, completeness is not sufficient for a 'well-behaved' representation, as is evidenced by the unwelcome behaviour of the Gabor representation at the critical density of one vector per unit cell [5]. To overcome this problem, the following definition is introduced.

Definition. A set $\mathcal{G}$ of vectors $g_n \in L^2(R)$ is said to be *locally bounded* in $L^2(R)$ if there exists a constant $A > 0$ such that if $[\xi_1, \xi_2)$ is any finite interval, then any $f \in L^2(\xi_1, \xi_2)$ can be written as $f = \sum_n \alpha_n g_n$, where the scalars α_n satisfy $\alpha_n = 0, g_n \notin \mathcal{G}(\xi_1, \xi_2)$, and $\sum_n |\alpha_n|^2 \leq A||f||^2$.

Obviously a locally bounded set is both bounded and complete. A less obvious consequence is contained in the following lemma.

Lemma. Let the set $\mathcal{G} = \{g_n, n \in Z_+\}$ be locally bounded in $L^2(R)$, with bound A, and let $\mathcal{G}_N = \mathcal{G} - \{g_n, n \leq N\}$ be the subset of $\mathcal{G}$ containing all but the first N vectors. Then $\mathcal{G}_N$ is also locally bounded with bound A.

The proof of this lemma is contained in the Appendix, along with that of the following theorem, which answers the question : " Can a locally bounded set be a frame ?".

Theorem. No set $\mathcal{G}$ which is locally bounded in $L^2(R)$ is a frame.

In other words, no frame is 'big enough' to be locally bounded. To construct a locally bounded set, it is convenient to start with a set of vectors with support $[0, 1)$, $\mathcal{F} = \{f_m, m \in Z\}$, say, and apply the standard wavelet construction [27],

$$g_{0m0} = f_m \tag{4}$$

and

$$g_{lmn}(\xi) = s^{n/2} g_{0m0}(s^n \xi - l) \tag{5}$$

This has the obvious merits of simplicity of construction and closure of the set $\mathcal{G} = \{g_{lmn}, l, m, n \in Z\}$ under the subgroup of dilations of the form $\xi \to s^n \xi$. This leaves open the choice of the set $\mathcal{F}$, of course.

Now the other significant requirement on the set $\mathcal{G}$ is some form of translation invariance : '*what* it is' should be as far as possible independent of '*where* it is' [12, 30]. Unfortunately, the set $\mathcal{G}$ is not even closed under a subgroup of translations. On the other hand, there are subsets of $\mathcal{G}$ which are closed, since for any l, m, n

$$g_{lmn}(\xi + s^{-n}k) = g_{(l+k)mn}(\xi) \qquad k \in Z \tag{6}$$

The problem is what to do with the 'residual' translations, modulo the interval of support. In other words, is it possible to choose $\mathcal{F}$ so that it is closed under the set of translations in the range $(-1, 1)$? It is immediately apparent that $\mathcal{F}$ cannot contain all such translations of a given vector f_m. Alternatively, it could be required that the vectors f_m, or rather the subspaces they define, be translation invariant, ie

$$f_m(\xi - \chi) = \lambda(\chi) f_m(\xi) \qquad -1 \leq \chi \leq 1 \tag{7}$$

where $|\lambda(\chi)| = 1$. Of course, no f_m with compact support is translation invariant : locality and translation invariance conflict in a way which relates directly to the uncertainty principle. It is, however, possible to reformulate the requirement in a less demanding way, using the resolution of the identity [23] associated with the translation group. To this end, consider the set of unit norm vectors $\mathcal{F} = \{f_m, m \in Z\}$, whose support is the unit interval $[0, 1)$, which maximise the functionals $I_m(f)$ defined by

$$I_m(f) = \int_0^1 d\xi \int_0^1 d\chi \rho_m(\xi - \chi) f^*(\xi) f(\chi) \tag{8}$$

where the kernel $\rho_m(\xi) = \exp[\jmath 2\pi m \xi] \sin[\pi\xi]/\pi\xi$ and $\jmath = \sqrt{-1}$. In effect, this replaces the ill-posed demand of (7) by the achievable one of maximising the correlation between the vector and its translations under the density defined by the kernel ρ_m.

The following assertions about this variational problem are well known or are easily proved :

1. The kernels ρ_m define projection operators P_m

$$P_m f(\xi) = \int_{-\infty}^{\infty} d\chi \rho_m(\xi - \chi) f(\chi) \tag{9}$$

where the set $\{P_m, m \in Z\}$ is a resolution of the identity [23] : $P_m P_n f = \delta_{mn} P_m f$ and $\sum_m P_m f = f, f \in L^2(R)$. The kernels ρ_m are just ideal bandlimiting filters with disjoint bands of width 2π radians [24, 25].

2. The vector f_m maximising $I_m(f)$ is the solution of the eigenvalue problem

$$\int_0^1 d\chi \rho_m(\xi - \chi) f_m(\chi) = \lambda f_m(\xi) \tag{10}$$

corresponding to the largest eigenvalue $\lambda > 0$ and $f_m(\xi) = \exp[\jmath 2\pi m\xi] f_0(\xi)$. The vector f_0 is the prolate spheroidal function with largest eigenvalue for the intervals $[0,1]$ and $[-\pi,\pi]$ [24]. Because they maximise the functionals $I_m(f)$, each f_m defines a subspace of $L^2(0,1)$ which is *minimally variant* to translations, under the density ρ_0 (which is positive on (-1,1)). In this sense, they satisfy the demand for translation invariance better than any other vectors in $L^2(0,1)$.

3. The set $\mathcal{F}$ is an exact frame for $L^2(0,1)$. Indeed, any $f \in L^2(0,1)$ can be written uniquely as $f = \sum_m \alpha_m f_m$ where $\alpha_m = (f, h_m)$ and the vectors $h_m(\xi) = f_0^{-1}(\xi)\exp[\jmath 2\pi m\xi]$ are bi-orthogonal to the f_m : $(f_m, h_n) = \delta_{mn}$.

4. Now define the vectors g_{lmn} by $g_{0m0} = f_m$ and $g_{lmn}(\xi) = s^{n/2} g_{0m0}(s^n\xi - l)$, where $s > 1$. Then the set $\mathcal{G} = \{g_{lmn}, l, m, n \in Z\}$ is locally bounded in $L^2(R)$. Defining $\tilde{g}_{lmn}$ in an analogous way, with $\tilde{g}_{0m0} = h_m$, then for each $n \in Z$, any $f \in L^2(R)$ can be written as $f = \sum_{l,m} \alpha_{lmn} g_{lmn}$, where the coefficient functionals $\alpha_{lmn} = \alpha_{lmn}(f) = (f, \tilde{g}_{lmn})$. The set of coefficient functionals $\{\alpha_{lmn}(f), l, m, n \in Z\}$ is called the MFT of f. While some minor modifications can be made to improve *snugness* [1] or simplify computation, the forms used below are all based directly on this construction.

5. Let $f \in L^2(\xi_1, \xi_2)$ be of compact support and let $f' \in L^2(\xi_3, \xi_4)$, where $\xi_3 > \xi_2$. Then there exists a scale n_0 and an index l_0 for which $\alpha_{lmn}(f + f') = \alpha_{lmn}(f), n \geq n_0, l \leq l_0$, or

$$f = \sum_{l \leq l_0} \sum_m \alpha_{lmn_0}(f + f') g_{lmn_0} \tag{11}$$

To see this, it suffices to choose n_0 such that $s^{-n_0} < \xi_3 - \xi_2$ and then choose l_0 so that $\xi_2 \leq s^{-n_0} l_0 < \xi_3$. The result follows immediately : the goal of providing an interference-free representation of arbitrary signals of compact support has been achieved.

The final matter to attend to is the selection of the *scale constant* s. In a WT, the scale constant is directly tied to the frame, but in the MFT it would appear to be somewhat arbitrary. Of course, the choice $s = 2$ has great attraction from a computational viewpoint, but this begs the question of whether there might be more compelling reasons for choosing a particular value of s. Since it has no connection with the completeness or boundedness of the set $\mathcal{G}$, the scale constant in the MFT can only be related to the problem of scale consistency - are some choices of s better suited than others to the comparison of inferences made at neighbouring scales ?

It is possible to address this issue in a way which is independent of the specific application by noting that the choice of s affects the relationship between the intervals of support of the vectors g_{lmn} in a set $\mathcal{G}$ of the form in 4. above. Consider rational forms $s = N/M$, where $M, N \;\; M < N$ are positive and mutually prime. Now consider the interval $[0, M)$. This is covered by M intervals of the form $I_{k0} = [k, k+1)$ which are the support of the vectors g_{km0} and N intervals of the form $I_{k1} = [kM/N, (k+1)M/N)$ which correspondingly support vectors g_{km1}. Now consider an arbitrary pair of such intervals I_{j0} and I_{k1}. They may be disjoint, in which case there is no reason to expect any agreement between the inferences made on the two, or they may have a non-empty intersection, in which case agreement is possible, but not guaranteed : only if they were identical could agreement be certain. Moreover, it seems reasonable to expect that the closer is the average intersection $I_{j0} \cap I_{k1}$ to the union $I_{j0} \cup I_{k1}$, the more likely is there to be agreement between inferences made at the two scales. To be more precise, the *consistency* of a set $\mathcal{G}$ of the form of 4. is defined below.

Definition. Let $\{I^s_{jk}, j, k \in Z\}$ be the intervals of support of the vectors $g^s_{kmj} \in \mathcal{G}^s$ defined in 4., with dependence on the scale constant $s > 1$ being made explicit. Let $\mu(A)$ be the measure of the subset $A \subset R$ and define $\nu(A)$ by

$$\nu(A) = \begin{cases} 0 & \text{if } \mu(A) = 0 \\ 1 & \text{else} \end{cases} \tag{12}$$

then the *consistency* of $\mathcal{G}^s$,$\rho(s)$, is defined to be

$$\rho(s) = \lim_{N \to \infty} \frac{\sum_{j=-N}^{N} \sum_k \mu(I^s_{j0} \cap I^s_{k1})}{\sum_{j=-N}^{N} \sum_k \mu(I^s_{j0} \cup I^s_{k1}) \nu(I^s_{j0} \cap I^s_{k1})} \tag{13}$$

Since $s > 1$, it follows immediately that $\rho(s) < 1$. In fact, the maximum that $\rho(s)$ can attain is given in the propostion below.

Proposition. No set $\mathcal{G}^s$ of the form in the definition can have a higher consistency than $1/2$. The maximum is attained only if the scale constant $s = 2$.

The proof is contained in the Appendix. Thus a scale constant $s = 2$ is not just attractive computationally - it also proves the best choice in terms of maximising the likelihood of agreement between inferences made on two overlapping intervals at adjacent scales. This greatly simplifies the modelling and inference processes upon which the segmentation is based, as is described in the references [18, 15, 17].

2.2 The continuous MFT

It is helpful at this point to define the continuous transform because it allows the principles behind the analysis methods to be demonstrated rather simply and is rather straightforward notationally. The primary aim of this section is to give the reader a feeling for the ways in which the inclusion of scale in Fourier analysis can be useful.

Although it is natural in this case also to use some form of prolate spheroidal function for the data window [22, 25](cf section 2.1), it suffices for the sequel that the window be of finite energy and both it and its Fourier Transform (FT) be smooth ($\mathcal{C}^2$) and even. The MFT of $x(\xi)$ at position ξ, frequency ω and scale σ is defined by

$$\hat{x}(\xi,\omega,\sigma) = \sigma^{1/2}\int_{-\infty}^{\infty} d\chi x(\chi) w(\sigma(\chi-\xi))\exp[-\jmath\omega\chi] \tag{14}$$

where $\jmath = \sqrt{-1}$.This differs from a conventional STFT only in the explicit reference to the scale of analysis : whereas a conventional phase-space description has $2m$ dimensions, for an $m-d$ signal, the MFT has $(2m+1)$. This has some analogy with the 'scale-space' methods developed by Witkin [20]. Assuming, as would be the case for any reasonable window, that

$$w(0) > 0 \tag{15}$$

and

$$\int_{-\infty}^{\infty} d\xi w(\xi) = A_w > 0 \tag{16}$$

then the original function can be regarded, at least formally, as the limit as $\sigma \to \infty$ of its MFT

$$x(\xi) = A_w^{-1}\lim_{\sigma\to\infty}\sigma^{1/2}\exp[\jmath\omega\xi]\hat{x}(\xi,\omega,\sigma) \tag{17}$$

and similarly, its FT $\hat{x}(\omega)$ as the limit

$$\hat{x}(\omega) = \lim_{\sigma\to 0}\sigma^{-1/2}\hat{x}(\xi,\omega,\sigma) \tag{18}$$

An alternative route to the MFT is via the FT $\hat{x}(\omega)$

$$\hat{x}(\xi,\omega,\sigma) = \frac{\sigma^{-\frac{1}{2}}}{2\pi}\int_{-\infty}^{\infty} d\rho\hat{x}(\rho)\hat{w}\left(\frac{\omega-\rho}{\sigma}\right)\exp[\jmath\xi(\rho-\omega)] \tag{19}$$

There are several ways to invert the MFT, as one might expect. These typically use a 'synthesis window' $v(\xi)$ with similar properties to the analysis window and such that

$$\int_{-\infty}^{\infty} d\xi w(\xi) v(-\xi) = 1 \tag{20}$$

With this choice, inversion exactly parallels the STFT, with the minor but useful generalisation of averaging over scale, with respect to a 'scale density' $p_\xi(\sigma)$, giving

$$x(\xi) = \frac{1}{2\pi}\int_0^{\infty} d\sigma p_\xi(\sigma)\sigma^{1/2}\int_{-\infty}^{\infty} d\chi\int_{-\infty}^{\infty} d\omega\hat{x}(\chi,\omega,\sigma)v(\sigma(\chi-\xi))\exp[\jmath\omega\xi] \tag{21}$$

where

$$\int_0^{\infty} d\sigma p_\xi(\sigma) = 1 \qquad -\infty < \xi < \infty \tag{22}$$

In particular, if $p_\xi(\sigma) = \delta(\sigma - a(\xi))$, $a(\xi) > 0$, $-\infty < \xi < \infty$, the scale of the transform is selected as a function of position. An alternative inversion uses only scale and frequency

$$x(\xi) = \frac{\xi}{\pi}\int_0^{\infty} d\sigma\sigma^{-1/2}\int_{-\infty}^{\infty} d\omega\hat{x}(0,\omega,\sigma)v(-\sigma\xi)\exp[\jmath\omega\xi] \tag{23}$$

as is readily verified by substitution from (14) into (23), using (20).

It was mentioned that the MFT is a 'superset' of the STFT and wavelet transforms. This is illustrated by noting the symmetry properties of the vectors $\tilde{g}_{\xi,\omega,\sigma}(\chi)$ [1],

$$\tilde{g}_{\xi,\omega,\sigma}(\chi) = w(\sigma(\chi - \xi)) \exp[-\jmath\omega\chi] \tag{24}$$

Translation of such a vector by δ gives the vector

$$w(\sigma(\chi + \delta - \xi)) \exp[-\jmath\omega(\chi + \delta)] = \exp[-\jmath\omega\delta]\tilde{g}_{\xi-\delta,\omega,\sigma}(\chi) \tag{25}$$

while a shift in frequency by ρ gives

$$w(\sigma(\chi - \xi)) \exp[-\jmath(\omega + \rho)\chi] = \tilde{g}_{\xi,\omega+\rho,\sigma}(\chi) \tag{26}$$

and a dilatation by a factor ϵ gives

$$w(\sigma(\epsilon\chi - \xi)) \exp[-\jmath\omega\epsilon\chi] = \tilde{g}_{\epsilon^{-1}\xi,\epsilon\omega,\epsilon\sigma}(\chi) \tag{27}$$

The vectors thus transform amongst themselves under the action of a group of symmetries which includes as subgroups the Weyl-Heisenberg and affine groups [1].

It is now appropriate to explore the ways in which the added complexity of the MFT, over that of its 'parents', can be used to advantage in signal analysis. The examples used in this section are intended to illustrate the general ideas behind its applications.

The first example concerns the modulated complex exponential signal

$$x(\xi) = a(\xi) \exp[\jmath\phi(\xi)] \tag{28}$$

for which it is assumed that both the amplitude modulation $a(\xi)$ and instantaneous frequency $\phi'(\xi)$ are smooth ($\mathcal{C}^2$) functions of ξ. The MFT of $x(\xi)$ is

$$\hat{x}(\xi, \omega, \sigma) = \sigma^{1/2} \int_{-\infty}^{\infty} d\chi a(\chi) w(\sigma(\chi - \xi)) \exp[\jmath(\phi(\chi) - \omega\chi)] \tag{29}$$

By change of variable, this becomes

$$\hat{x}(\xi, \omega, \sigma) = \sigma^{-1/2} \int_{-\infty}^{\infty} d\chi a(\xi + \frac{\chi}{\sigma}) w(\chi) \exp[\jmath(\phi(\xi + \frac{\chi}{\sigma}) - \omega(\xi + \frac{\chi}{\sigma}))] \tag{30}$$

which, for large σ, can be approximated by

$$\hat{x}(\xi, \omega, \sigma) = \sigma^{-1/2} \int_{-\infty}^{\infty} d\chi a(\xi) w(\chi) \exp[\jmath(\phi(\xi) + \frac{\chi}{\sigma}\phi'(\xi) - \omega(\xi + \frac{\chi}{\sigma}))] + O(\sigma^{-2}) \tag{31}$$

or

$$\hat{x}(\xi, \omega, \sigma) \approx \sigma^{-1/2} a(\xi) \hat{w}\left(\frac{\omega - \phi'(\xi)}{\sigma}\right) \exp[\jmath(\phi(\xi) - \omega\xi)] \tag{32}$$

Thus $\hat{x}(\xi, \omega, \sigma)$ has an envelope which is a product of the FT of the scaled window, centred on the instantaneous frequency $\phi'(\xi)$, and the AM envelope $a(\xi)$. By differentiating its phase, the instantaneous frequency can be recovered

$$\phi'(\xi) = \frac{\partial \arg \hat{x}}{\partial \xi} + \omega \tag{33}$$

It can also be seen that the magnitude of the MFT is significant in a region of the $\xi - \omega$ plane determined by the signal duration and frequency, the spread in frequency being dependent on the bandwidth of the modulation, as common sense would suggest.

When several such signals, $x_i(\xi), 1 \leq i \leq I$, are present, they can be separated, provided that their instantaneous frequencies are sufficiently far apart, ie if

$$\hat{w}\left(\frac{\phi_i'(\xi) - \phi_j'(\xi)}{\sigma}\right) \ll \hat{w}(0) \qquad i \neq j \tag{34}$$

for some $\sigma > 0$ such that the approximation (32) is valid. In other words, in the cases of practical interest, where segmentation of the individual signals $x_i(\xi)$ is required, there will be both an upper bound $\sigma_u(\xi, \omega)$ on the scale, set by the local frequency separation of the signals, and a lower bound $\sigma_l(\xi, \omega)$, set by their local modulation rate. This form of signal is rather typical of those encountered in the analysis of polyphonic music.

It may be instructive to compare the MFT with the STFT and conventional wavelet analysis for such signals. Of course, if there exists a fixed scale σ which always lies between the upper and lower bounds, there is no problem in using an STFT. Unfortunately, this appears not to be the case in general - the scale of analysis has to be varied with both frequency and position, in a way which cannot be predicted a priori. The wavelet transform of the signal is readily found from (32) by setting $\omega = 0$ and noting that the window must be replaced by a suitable wavelet, which requires[1]

$$\int_{-\infty}^{\infty} d\xi w(\xi) = 0 \tag{35}$$

For this transform, the envelope becomes

$$|\hat{x}(\xi, 0, \sigma)| = \sigma^{-1/2}|a(\xi)| \left|\hat{w}\left(\frac{-\phi'(\xi)}{\sigma}\right)\right| \tag{36}$$

To have significant amplitude, the scaled instantaneous frequency $\phi'(\xi)$ must lie within the frequency band $[\omega_l, \omega_u]$ for which the FT of the wavelet differs significantly from zero. If two such signals are present, with $\phi_1'(\xi) > \phi_2'(\xi)$, this requires the existence of scales $\sigma_1 > \sigma_2$ such that

$$\omega_l < \frac{\phi_i'}{\sigma_i} < \omega_u \qquad i = 1, 2 \tag{37}$$

but

$$\frac{\phi_1'}{\sigma_2} > \omega_u \qquad \frac{\phi_2'}{\sigma_1} < \omega_l \tag{38}$$

This imposes additional constraints on the scale, which are unlikely to be achieved in practice. In effect, the scale is tied to the carrier frequency of the signal, rather than its modulation frequency. To put it another way, in a conventional wavelet transform, scale is an integral part of the phase-space description - it simply replaces frequency as the conjoined co-ordinate. In the MFT, it is a free parameter, which can be chosen to simplify the basic phase-space description.

A second useful example, in many ways the dual of the first, is that of a transient signal. There are many possible definitions of such signals, but their main features are (i) a discontinuity and (ii) rapid decay away from that discontinuity. An appropriate choice, particularly for the image analysis application, is

$$x(\xi) = \begin{cases} -0.5x_0 \exp[-\alpha(\xi - \xi_0)] & \text{if } \xi > \xi_0 \\ 0.5x_0 \exp[\alpha(\xi - \xi_0)] & \text{if } \xi < \xi_0 \end{cases} \tag{39}$$

In this case, the MFT is conveniently defined through the FT $\hat{x}(\omega)$,

$$\hat{x}(\omega) = \frac{\jmath\omega x_0 \exp[-\jmath\omega\xi_0]}{\omega^2 + \alpha^2} \tag{40}$$

giving, from (19)

$$\hat{x}(\xi,\omega,\sigma) = \frac{\sigma^{-\frac{1}{2}}}{2\pi} \int_{-\infty}^{\infty} d\rho \frac{\jmath\rho x_0}{\rho^2 + \alpha^2} \hat{w}\left(\frac{\omega - \rho}{\sigma}\right) \exp[\jmath\rho(\xi - \xi_0) - \omega\xi)] \tag{41}$$

which, on change of variable and choosing σ small, can be approximated by

$$\hat{x}(\xi,\omega,\sigma) \approx \sigma^{1/2} w(\sigma(\xi_0 - \xi)) \frac{\jmath\omega x_0 \exp[-\jmath\omega\xi_0]}{\omega^2 + \alpha^2} \tag{42}$$

which has an envelope which is the product of the window, centred at ξ_0, and the FT of the signal, $\hat{x}(\omega)$. The position of the discontinuity, ξ_0, can be recovered exactly by differentiating the phase of $\hat{x}(\xi,\omega,\sigma)$

$$\xi_0 = -\frac{\partial \arg \hat{x}}{\partial \omega} \qquad \omega \neq 0 \tag{43}$$

As in the exponential signal, the MFT at the appropriate scale seems to express the obvious signal features in a natural way. If there are several such signals, they may still be separable over some range of scales $\sigma_l(\xi) < \sigma < \sigma_u(\xi)$. Although no direct comparison with wavelet methods is possible in this case, the use of zero-crossing analysis of luminance discontinuities in images, based on scale-space or wavelet descriptions, is similar and has been widely reported in the literature [20, 21, 31]. In the analysis of musical signals, the use of transient analysis is limited by its dependence on the instrument playing the music - percussive instruments often produce some transient energy, but this generally lacks the essential ingredient of a discontinuity.

2.3 The MFT in higher dimensions

As noted above, the MFT of an $m-d$ signal is a $(2m+1)-d$ object, expressible as the obvious extension to (1) for a signal $x(\vec{\xi})$ which is a scalar function of the $m-d$ co-ordinate $\vec{\xi} = (\xi_1, ..., \xi_m)^T$

$$\hat{x}(\vec{\xi}, \vec{\omega}, \sigma) = \sigma^{m/2} \int_{-\infty}^{\infty} d\vec{\chi} x(\vec{\chi}) w(\sigma(\vec{\chi} - \vec{\xi})) \exp[-\jmath \vec{\omega}.\vec{\chi}] \tag{44}$$

where $\vec{\omega}$ is the Fourier co-ordinate and '.' denotes scalar product.

If the window $w(\vec{\xi})$ is isotropic, ie if $w(\vec{\xi}) = w(||\vec{\xi}||)$, then the $m-d$ MFT has a rather nice symmetry property. To illustrate this point, let T be an affine transform of the form

$$T\vec{\xi} = \sigma_0 R\vec{\xi} + \gamma_0 \tag{45}$$

where R is a rotation, then the MFT of the transformed signal $Tx(\vec{\xi}) = x(T^{-1}\vec{\xi})$ [28] is given, after substitution from (45), by

$$\sigma^{m/2} \int_{-\infty}^{\infty} d\vec{\chi} x(T^{-1}\vec{\chi}) w(\sigma(\vec{\chi} - \vec{\xi})) \exp[-\jmath \vec{\omega}.\vec{\chi}] = \hat{x}(T^{-1}\vec{\xi}, \sigma_0 R^{-1}\vec{\omega}, \sigma\sigma_0) \exp[-\jmath \vec{\omega}.\vec{\gamma_0}] \tag{46}$$

In other words, the MFT of the transformed signal can be found directly from that of the original signal, simply by an appropriate change of co-ordinates and phase shift. Correspondingly, the set of MFT vectors is closed under the group of co-ordinate transformations of the form of (45).

The last example of this section, which is related to the above symmetry property, is chosen to illustrate the use of scale in multidimensional analysis in a way that is particularly relevant to the image analysis work. It is based on the idea of 'local one-dimensionality' [32, 33] : many signals in $m-d$ spaces can be approximated in a neighbourhood of each point $\vec{\xi}$ by a function of $1-d$ variation. In such a case, for any vector $\vec{\chi}$, there is some $\delta > 0$ such that

$$x(\vec{\xi} + \epsilon\vec{\chi}) = x_1(\epsilon\vec{\chi}.\vec{\theta}(\vec{\xi})) + O(\epsilon^2) \qquad \forall \vec{\xi}, \;\; |\epsilon| < \delta \tag{47}$$

where $\vec{\theta}(\vec{\xi})$ is a unit vector, effectively a generalisation of the gradient, since the approximation (47) may be used in some cases where $x(\vec{\xi})$ is discontinuous, such as at luminance discontinuities in images.

Substitution from (47) into (44) gives

$$\hat{x}(\vec{\xi}, \vec{\omega}, \sigma) \approx \exp[-\jmath \vec{\omega}.\vec{\xi}] \int_{-\infty}^{\infty} d\vec{\chi} x_1(\sigma^{-1}\vec{\chi}.\vec{\theta}(\vec{\xi})) w(\vec{\chi}) \exp[-\jmath \sigma^{-1} \vec{\omega}.\vec{\chi}] \tag{48}$$

To make further progress, it is easiest to assume a Gaussian form of window,

$$w(\vec{\xi}) = \exp[-0.5\vec{\xi}.\vec{\xi}] = \prod_{i=1}^{m} w_1(\xi_i) \tag{49}$$

where $w_1(\xi)$ is the $1-d$ window. This has the useful features of being both rotation-invariant and cartesian separable. Taking advantage of these allows a rotation of axes and separation, giving

$$\hat{x}(\vec{\xi},\vec{\omega},\sigma) \approx \exp[-\jmath\vec{\omega}.\vec{\chi}]\hat{x}_1(0,\omega_1(\vec{\theta}(\vec{\xi})),\sigma)\prod_{i=2}^{m}\hat{w}_1\left(\frac{\omega_i(\vec{\theta}(\vec{\xi}))}{\sigma}\right) \tag{50}$$

where $\omega_i(\vec{\theta}(\vec{\xi}))$ is the ith rotated Fourier co-ordinate

$$\omega_1(\vec{\theta}(\vec{\xi})) = \vec{\theta}(\vec{\xi}).\vec{\omega} \tag{51}$$

In other words, the MFT has its energy concentrated along the direction of variation of the signal, with a bandwidth about that line which depends on the window scale required to make the approximation valid. This in turn depends upon the rate of change of the direction $\vec{\theta}(\vec{\xi})$ with position - the curvature of the neighbourhood around $\vec{\xi}$. Thus in higher dimensions, the scale can be varied to estimate the local geometry implied by the signal, an important task in a number of image analysis applications. The conclusion remains valid, even when the window is not Gaussian - it is simply messier to prove it.

2.4 The discrete MFT

In analogy with the continuous transform, the $1-d$ MFT component at scale $\sigma(n)$, position $\xi_i(n)$ and frequency $\omega_j(n)$ can be written as

$$\hat{x}(\xi_i(n),\omega_j(n),\sigma(n)) = \sum_k w_n(\xi_k - \xi_i(n))x(\xi_k)\exp\left[-\jmath\xi_k\omega_j(n)\right] \tag{52}$$

where $w_n(\xi_k)$ is the scale n window sequence, ξ_k is the kth sample point for the original sequence and the sample points at scale n are $\xi_k(n)$ and $\omega_l(n)$ for the original and Fourier co-ordinates respectively. The window sequences at different scales are approximately related through a continuous window function $w_c(\xi)$

$$w_n(\xi_k) \approx w_c(\sigma^n\xi_k) \quad n > 0 \tag{53}$$

where $\sigma > 1$ is the *scale constant* of the MFT. In the applications, $\sigma = 2$ and the data set is of finite extent $N = 2^M$. This allows efficient computation by means of an algorithm based on the FFT. The general structure of the MFT, as a function of the scale index n, is shown for a simple $2-d$ case in Fig. 1. If the notation for the continuous MFT is clumsy, that for the discrete case is worse. It is often convenient to hide the details of sampling in the vector notation

$$\hat{\boldsymbol{x}}(n) = \boldsymbol{F}(n)\boldsymbol{x} \tag{54}$$

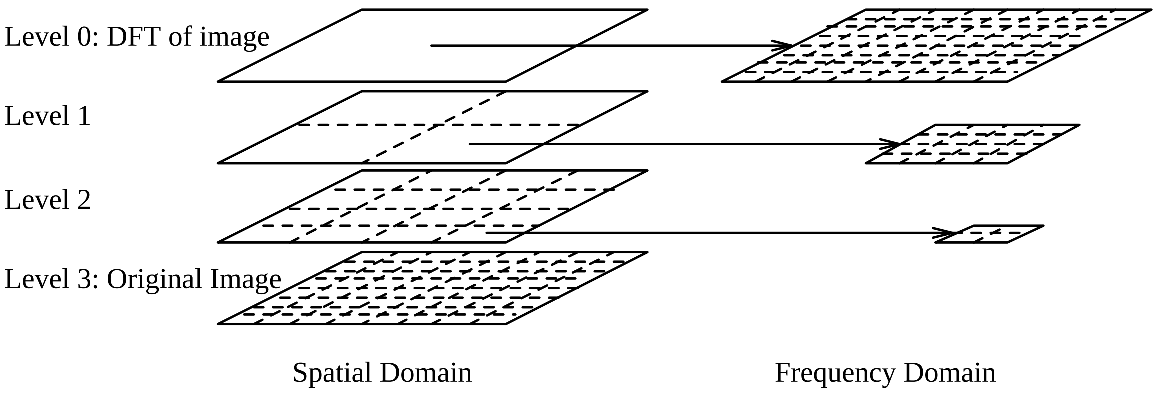

Figure 1: MFT Structure

in which $\hat{\boldsymbol{x}}(n)$ and $\boldsymbol{x}$ are vectors whose components $\hat{x}_{ij}(n)$ and x_k are related by the nth level transform operator $\boldsymbol{F}(n)$. It is also useful to consider, if the data set is finite, the unwindowed transform operator $\boldsymbol{F}(0) = \boldsymbol{F}$, for which

$$w_0(\xi_k) = N^{-1/2} \quad 0 \leq k < N \tag{55}$$

and to include, at the other extreme of scale, the window

$$w_M(\xi_k) = \delta_{0k} \tag{56}$$

the unit impulse sequence, which corresponds to the trivial transform operator $\boldsymbol{F}(M) = \boldsymbol{I}$, the identity.

As in the continuous case, the discrete MFT is invertible from any level of the representation. In other words, for each n there is an inverse operator $\boldsymbol{F}^{-1}(n)$ such that

$$\boldsymbol{x} = \boldsymbol{F}^{-1}(n)\hat{\boldsymbol{x}}(n) \tag{57}$$

The need for invertibility imposes strict requirements on the sampling intervals and windows, which are discussed below.

The key factors determining the usefulness of the MFT are the distribution of sample points in the signal and frequency co-ordinates and the related problem of selecting a good window (cf section 2.1). As has been noted in a number of places, these are closely related by the uncertainty principle (eg [1, 7, 12, 24, 22]). Starting from a regular (periodic) lattice in $m{-}d$ containing N^m points, common sense suggests that for a complete representation, the number of sample points along each axis in the ξ-co-ordinates, $N_\xi(n)$, and in the ω-co-ordinates, $N_\omega(n)$ should satisfy the inequality

$$N_\xi(n)N_\omega(n) \geq N \tag{58}$$

To satisfy the invertibility requirement of (57), this means that if equality holds in relation (58), the vectors forming the rows of $\boldsymbol{F}(n)$ must be linearly independent - they constitute a basis. If the inequality holds, then the set of vectors will be called a frame, in keeping with the terminology of [1]. As in the STFT, the sample points at level n of the MFT are spaced at regular intervals in both the ξ-co-ordinates and the ω-co-ordinates. Assuming that $N = 2^M$, then an obvious choice for the sampling intervals $\Xi(n)$ and $\Omega(n)$ is

$$\Xi_k(n) = 2^{M-k-n} \quad \Omega(n) = 2^{n+1-M}\pi \tag{59}$$

so that

$$N_\xi(n) = 2^{n+k} \quad N_\omega(n) = 2^{M-n} \tag{60}$$

and

$$N_\xi(n)N_\omega(n) = 2^k N \tag{61}$$

which implies an oversampling by a factor 2^k. The reason for introducing this redundancy is partly explained by the arguments in [1] : oversampling is a necessary price for using frames that are both well-behaved numerically, or *snug*, and *localised* in some appropriate sense in both the original and frequency co-ordinate systems.

Questions of snugness and locality are essentially bound up with the choice of window : a good choice of window should result in good localisation, efficient computation and snugness. An obvious choice candidate in a general tool for Fourier analysis is an appropriate form of prolate spheroidal sequence [13, 25]. Indeed, one form of these was used to good effect in one typical $2-d$ application of spectrum estimation - texture segmentation [7]. In that application, a non-redundant form of wavelet transform was used. It was shown in [7], (77),(78), that the transform was snug, using a definition borrowed from [26], which is closely related to the *frame bounds* A and B defined by Daubechies in [1]. It was shown in that work that efficient computation of the transform could be obtained if the window sequence is bandlimited, using FFT's. The total computation in that case was shown to be equivalent to an FFT-based convolution operation, ie of order $N^2 \log_2 N$ for an image of size $N \times N$.

In more general applications, however, and particularly if inversion of the transform is required, then a significant issue is the locality not only of the analysis window $\boldsymbol{w}_n$ but also of the corresponding synthesis window $\boldsymbol{v}_n$ of the inverse transform operator (corresponding to the functions $\tilde{g}_{mn}$ of [1]). Simple-minded application of the appropriate prolate spheroidal equation will result, if $k = 0$, in a window whose inverse is a good deal less well localised than an ideal bandlimiting filter - hardly a good choice.

A second problem relates to the definition and measurement of position in the original and Fourier co-ordinates was discussed at some length above, for the continuous MFT. In the discrete case, however, the partial derivatives of (33), (43) must be approximated by differences, using, for example,

$$\tau_{ij}(n) = \arg \hat{x}_{ij}(n)\hat{x}^*_{i(j-1)}(n) \tag{62}$$

and

$$\nu_{ij}(n) = \arg \hat{x}_{ij}(n)\hat{x}^*_{(i-1)j}(n) \tag{63}$$

While such definitions are always possible, their interpretation is unclear unless signals which have unambiguous positions in one or other domain are chosen. These are, of course, the vectors defining the irreducible representations of the corresponding groups of translations [28]. Consider, therefore, an impulse at co-ordinate ξ_i, the sequence $\{x_k\} = \{\delta_{ik}\}$, with MFT components at $\xi_0 = 0$ giving, from (52), (62),

$$\tau_{0j}(n) = \xi_i\Omega(n) \quad \text{if} \;\; w_n(\xi_{-i}) \neq 0 \tag{64}$$

The *rhs* of (64) is, however, a phase : it only has a unique definition if $|\xi_i\Omega(n)| < \pi$. In other words, to avoid ambiguity, the window must satisfy, from (52),(63),

$$w_n(\xi_{-i}) = 0 \quad \text{if} \quad |\xi_i| > 2^{k-1}\Xi_k(n) \tag{65}$$

Now consider a signal with unambiguous frequency, a complex exponential with frequency ν. By similar reasoning, it can be concluded that the unambiguous definition of instantaneous frequency requires a window with FT $\hat{w}_n(\nu)$ satisfying

$$\hat{w}_n(\nu) = 0 \quad \text{if} \quad |\nu| > 2^{k-1}\Omega(n) \tag{66}$$

Of course, no such window sequence exists [1, 24]. The best that can be done is to concentrate the energy of the window inside the *unit cell* in the $\xi - \omega$ plane. This is a specification of the prolate spheroidal sequences and their relatives [7, 24, 13, 25], which may be defined as solutions of the eigenvalue problem

$$\boldsymbol{I}(\Xi)\boldsymbol{B}(\Omega)\boldsymbol{I}(\Xi)\boldsymbol{v} = \lambda\boldsymbol{v} \tag{67}$$

where $\boldsymbol{I}(\Xi)$ is the index-limiting operator

$$I_{ij}(\Xi) = \begin{cases} \delta_{ij} & |\xi_i| < \Xi/2 \\ 0 & \text{else} \end{cases} \tag{68}$$

and $\boldsymbol{B}(\Omega)$ is the bandlimiting operator

$$\boldsymbol{B}(\Omega) = \boldsymbol{F}^*\boldsymbol{I}(\Omega)\boldsymbol{F} \tag{69}$$

The vector $\boldsymbol{v}$ is index-limited, of course. The related bandlimited sequence $\boldsymbol{u}$ can be defined analogously by

$$\boldsymbol{B}(\Omega)\boldsymbol{I}(\Xi)\boldsymbol{B}(\Omega)\boldsymbol{u} = \lambda\boldsymbol{u} \tag{70}$$

The properties of such sequences have been rather fully explored [9, 24, 13].

In the present context, the bandlimited form has the computational advantages mentioned above and so attention will be focused on that form. When $k = 0$, the obvious choice of intervals at level n is $\Xi = \Xi_0(n)$ and $\Omega = \Omega(n)$, giving a reasonably snug frame and intervals which match the limits for the window in (65),(66). If k is increased, the situation is less clear. On one hand, for maximum localisation in terms of energy in the unit cell, it would clearly be advantageous to retain the same intervals. On the other hand, the reduction of the sampling interval Ξ_k implies that the bandwidth Ω can be increased by a factor of 2^k without violating the frequency ambiguity constraint, (66), and will give better localisation in the ξ-co-ordinate, leading to reduced ambiguity. It can also be expected to improve the snugness of the frame. Combining both of these demands gives the modified eigenvalue problem

$$\boldsymbol{B}(2^k\Omega(n))\boldsymbol{I}(\Xi_0(n))\boldsymbol{B}(\Omega(n))\boldsymbol{w}_n = \lambda'\boldsymbol{w}_n \tag{71}$$

The bandlimited sequence $\boldsymbol{w}_n$ is related to the original sequence $\boldsymbol{u}_n$(using $\Omega = \Omega(n)$ and $\Xi = \Xi_0(n)$ in (70)) by

$$\boldsymbol{w}_n = \boldsymbol{B}(2^k\Omega(n))\boldsymbol{I}(\Xi_0(n))\boldsymbol{u}_n \tag{72}$$

as is readily verified by premultiplying both sides of (70) by $\boldsymbol{B}(2^k\Omega(n))\boldsymbol{I}(\Xi_0(n)$ and using the relations $\boldsymbol{B}^2(\Omega) = \boldsymbol{B}(\Omega)$ and $\boldsymbol{B}(\Omega)\boldsymbol{u} = \boldsymbol{u}$. One way of describing the relation between the two sequences is that the bandwidth constraint is 'relaxed' in the definition of $\boldsymbol{w}_n$ compared with the original window.

The sequence $\boldsymbol{w}_n$ is thus that sequence which is bandlimited to an interval of length $2^k\Omega(n)$ and has maximum energy concentration in the unit cell defined by $\Xi_0(n)$ and $\Omega(n)$. In applications, the choice $k = 1$ has been found to give an adequate compromise between the increase in computation (a factor of 2 over the 'unrelaxed' window) and the benefits in terms of energy concentration, snugness and reduced ambiguity in estimating times via (64). In both cases, the resulting window has a FT $\hat{w}_n$ which is non-negative. In neither case is there ambiguity in the instantaneous frequency because the window is bandlimited in a way which satisfies the constraint equation (66). The ambiguity in the time estimate can be expressed in terms of the ratio of the largest sidelobe peak outside the constraint interval Ξ_0 to the peak magnitude in the window. For the case $k = 0$, this is -15.6dB, while for the relaxed case, it is -26.3dB, a significant improvement. The ratios of the *frame bounds* [1] also show the advantage of the choice $k = 1$: for $k = 0$, the ratio is typically 2.0, while for $k = 1$, it is 1.1. The two cases are illustrated in the time-frequency plots of Figs. 2 and 3, where the superior localisation of the relaxed window is self-evident. It turns out that the Fourier transform of $\boldsymbol{w}_n$, $\hat{\boldsymbol{w}}_n$ is quite close to the positive half-cycle of a cosine function - a generalised Hamming window in frequency [29]. Thus an easily

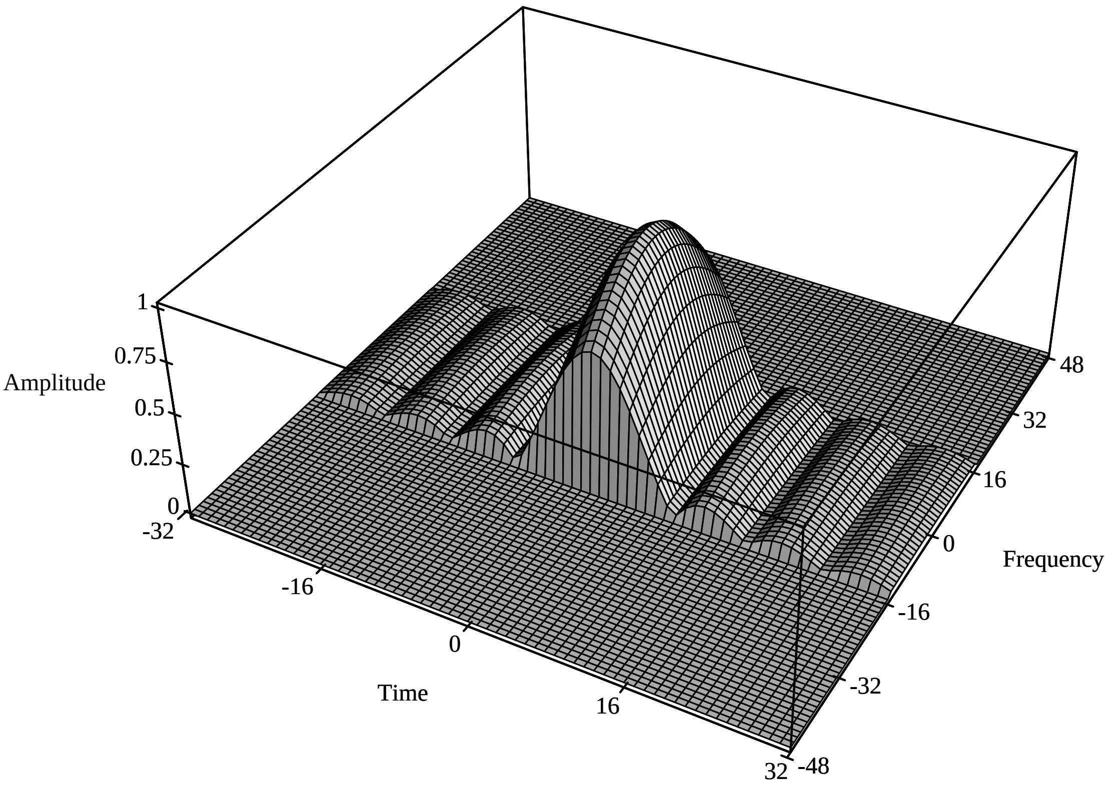

Figure 2: Time-Frequency plane FPSS 16×32

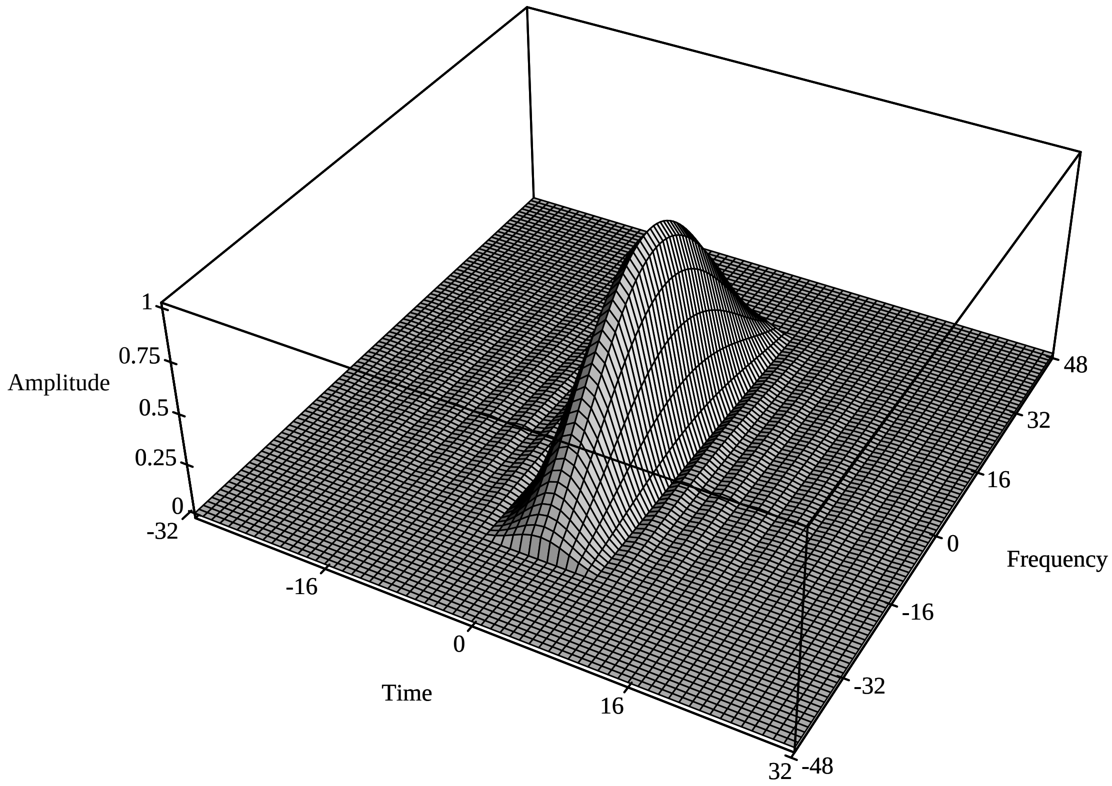

Figure 3: Time-Frequency plane 'relaxed' FPSS 16×32

computed 'sub-optimum' window is

$$\hat{w}'_n(\omega_k) = \begin{cases} \cos[\pi\omega_k/2\Omega(n)] & |\omega_k| \leq \Omega(n) \\ 0 & \text{else} \end{cases} \tag{73}$$

This choice has the interesting property of having the same synthesis window and hence of defining a tight frame [5]. The only cost associated with the choice is a slight increase in the sidelobe magnitudes.

Although the work described below is based exclusively on bandlimited windows, this does not imply that no corresponding decimation/interpolation filtering technique exists. One possibility for this would be to choose a set of quadrature-mirror filters by maximising energy concentration, subject to the 'QMF' constraint equations [27].

The final matter to attend to is the choice of the $m-d$ co-ordinates $\vec{\xi}_i(n)$ and $\vec{\omega}_j(n)$ for each level of the MFT. Because the transform algorithm is adapted to deal with data sets of size 2^{mM}, it is natural to space the co-ordinates at the centres of the corresponding 'blocks' of 2^{mn} samples. This means that if the original sample points are given by

$$\vec{\xi}_i = (i_1, i_2, ..., i_m) \quad \text{where} \quad 0 \leq i_l < 2^M \tag{74}$$

then

$$\xi_{il}(n) = \Xi_k(n)(i_l + 1/2) \qquad 1 \leq k \leq m \tag{75}$$

and analogously in the Fourier domain

$$\omega_{jl}(n) = \Omega_k(n)(j_l + 1/2) \qquad 1 \leq k \leq m \tag{76}$$

If the signal is not confined to a finite set of points, then an *overlap-add* technique can be used [29]. The window used in $m-d$ transforms has been chosen for simplicity to be the m-fold cartesian (tensor) product of the $1-d$ window $\boldsymbol{w}_n$

$$w_n^m(\vec{\xi}_i(n)) = \prod_{l=1}^{m} w_n(\xi_{il}) \tag{77}$$

There may well be applications where a different form of $m-d$ window is appropriate, for symmetry reasons, for example. This has not been found necessary in the image processing applications described below.

3 Applications

The results presented here are a brief summary of work described in greater detail in the references ([18, 15, 17, 19]). The aim of this section is to illustrate the versatility of the MFT as a tool for signal analysis.

3.1 Audio signal analysis

The results described briefly here are part of the experimental work completed by Pearson [17]. They show that by combining the results of note identification procedures from an appropriate range of scales, it is possible to achieve reliable segmentation of polyphonic music, in which an arbitrary number of voices may be playing simultaneously.

The first example consists of two isolated piano notes, $F\sharp 4$ and $C4$, which were taken from CD and spliced together so that there is noticeable time overlap between the two. Figs. 4 and 5 show the MFT's of this signal at two scales, the first corresponding to 187.5 time samples/sec and 21.33 frequency samples/kHz, while the second has 4 times the frequency resolution and a quarter of the time resolution. The transient energy caused by the percussive nature of the piano is clearly visible at the higher time resolution, but, on the other hand, at this scale the frequency resolution is inadequate to separate the *partials* (harmonics) associated with the two notes. It can be seen quite clearly from Fig 5 that although the partials of a given note occur at multiples of the fundamental frequency, suggesting perhaps a WT, this is no longer the case for a pair of notes. In this example, the closest pair of partials are the 2nd of $F\sharp 4$ and the 3rd of $C4$ - the separation being 45 Hz, significantly less than the fundamental frequency of either note. Thus even in this simple example, the limitations of WT methods in note segmentation are evident, while the MFT appears to be capable of dealing with the problem.

The more complex example shown in Figs. 6 and 7 is a 5 sec segment from the woodwind trio of Bach's 1st Brandenburg Concerto. In this case, the higher time resolution provides very limited information - the banding which is evident in many places is due to interference between nearby partials from different notes. At the higher frequency resolution, however, many more partials are individually identifiable. It may be worth noting that even in this short segment there are roughly 50 partials belonging to 17 notes played by the three instruments. There is some evidence of vibrato, particularly in the higher harmonics, but little sign of transient energy. There is also clear overlap in time between notes starting on different beats. The dynamic range spanned by the partials is also considerable and the detection problem is complicated by the presence of noise, rumble and inherent ambiguities, such as two instruments playing notes an octave apart on the third beat in the sequence. In summary, analysis of this signal is nontrivial.

Rather than going through the various stages of the segmentation process, which are described fully in [17], it suffices here to show the results of Figs. 8 and 9, which display the onsets, frequencies and tracks of the notes identified in the two examples. The algorithms are based on a hierarchical Markov model of the signal, which is similar to the Hidden Markov models used in speech recognition [40] and which, at the lowest level, are directly related to the example of a modulated exponential wave discussed in section 2. Detection of individual partials at various scales is followed by a scale consistency check, with identification of notes being the last step in the processing. Partials associated with various notes are labelled by the corresponding harmonic numbers. The two paino notes are succesfully identified, along with any partial having a magnitude up to 40dB below the largest one. In the segment of the Brandenburg Concerto, all 17 notes have been identified and their onset times estimated with an accuracy of about 0.1 sec. In addition, there were 5 'false alarms', mostly caused by second harmonics being labelled as notes. Many of these could be eliminated rather simply, but only at the cost of not identifying the two notes an octave apart. Ambiguities of this kind abound in music. To deal with them adequately, more sophisticated signal modelling, possibly taking harmonic and rhythmic structure into account, would be needed. Nonetheless, the result seems to compare well with those reported by others working in the field (eg [37, 38, 39]).

3.2 Image analysis

The use of Fourier analysis in image processing is perhaps less widely accepted than in music analysis. Recent years, however, have seen an upsurge in activity based on the use of the Gabor representation, or modifications of it which are perhaps more accurately described as WT's, for the simple reason that they employ frame vectors which are related by dilations (eg [12, 3, 8, 46, 45]). So far, such transforms have seen use in areas such as image data compression and texture analysis, both seen as potential applications for Fourier techniques for many years (eg [44, 43]). A less obvious, but important application, is in the extraction of boundary contours, a problem which was first tackled by applying a line following algorithm to the output of an edge detection process (eg [48, 49]). More recently, the line-following stage has been replaced by more effective, if expensive techniques such as some form of relaxation or energy minimization (eg [50, 51]), multiresolution methods [52, 53] or a parametric technique[54, 55]. A common feature of these methods is a reliance on some form of edge detection process, based on linear filtering. It is in this respect that the present work differs fundamentally from these methods : the edge and line extraction processes used in the present results employ a Markov model of the features operating directly on the MFT coefficients, based on the combination of the transient model and local one-dimensionality discussed in section 2. The resulting segmentation is fast, robust and capable of generalization [15, 19].

Two scales of the MFT of the well known 'Lena' image of Fig. 10 are shown in Figs. 11 and 12. At the higher spatial resolution, the local one-dimensionality of the edge features is clearly visible, with energy in the frequency domain being concentrated around a line in the direction of the normal to the curve. By modelling the phase and amplitude variation along these directions as a Markov process, it is possible to estimate the position and magnitude of the edge or line. These estimates can then be checked for scale consistency and combined to form the composite multilevel tessellation shown in Fig. 13. From this representation, it is then an easy job to link the contour segments into connected curves, as shown in Fig. 14. The total processing requirement for this image is of the order of 400 multiplications/pixel, equivalent to convolution with a pair of 14×14 spatial edge detection filters. Each continuous curve in Fig. 14 is stored as a single 'object' in the final representation.

This approach can be generalized quite easily to accommodate more complex feature models. Thus the results in Figs. 15-17 show the use of the same principles, but using feature models which include multiple straight lines and circle segments [62]. This is well illustrated in the synthetic 'Shapes' image of Figs. 15,16. Fig. 15 shows the edge features which have been extracted at varying scales, as indicated by the blocks whose edges aer shown at a low intensity. All of the vertices except those with accute angles have been located and the edge segments identified at differing scales have been linked into single features. The accuracy of the estimates is demonstrated in Fig 16, which shows the edge estimates highlighted on the original image. In Fig 17, a low resolution natural image has been used as the test data. Although some of the weaker features have been missed, the algorithm has performed adequately on the whole. Again, the 'primitive' features are modelled and estimated directly on the MFT coefficients, requiring a few hundred multiplications/pixel for the full detection procedure. Unlike the previous example, such features as sharp corners and junction points have been extracted directly from the MFT data. Thus the approach to segmentation via the MFT is general and computationally efficient.

4 Conclusions

The MFT has been presented as a way of combining STFT and wavelet methods into a single transform, which was shown in section 2 to be uniquely suited to the local analysis needed for segmentation. Whether it is seen as a systematic approach to window adaptation in STFT analysis [11] or as an enrichment of the WT is a matter of taste, in the authors' opinion. A more important question is whether it is an effective tool for analysing the complicated signals which seem both to defeat simpler analysis methods and to be rather common in nature. In this respect, the authors have been encouraged by the results of applying it to problems which have proved to be difficult to solve using conventional techniques. The results compare well with those reported elsewhere (eg [36, 37, 38, 39, 50, 51, 55]) and have been achieved at reasonable computational cost. Moreover, the signal models used in this work are themselves somewhat novel and appear to reflect the complex structures of 'real world' signals rather better than the simple statistical models which have been used in much previous work in the two fields.

In considering extensions of the work, one obvious area for attention is the theoretical effort in pinning down the notion of *locality* more adequately than was done in section 2 - the assumption that the analysis vectors must be of compact support could surely be replaced by one requiring that they have all but a small fraction of their energy in some finite interval, but this issue of 'ϵ-locality' has not been addressed. The effects of other symmetries in higher dimensions, such as rotations, have not been considered fully, although they are known to be important in a number of image analysis applications (eg [33, 58, 32, 59]). Nor can it be claimed that all the problems in the two application areas have been solved simply by use of the MFT. There are obvious limitations in the methods reported here, some of which have already been discussed. In the music segmentation, the need to incorporate musical knowledge in modelling the production of notes is clear from the results. A less obvious problem is that when many voices contribute energy at the same frequency, as occurs frequently in ensemble playing, the phase coherence which was used in the estimation of partial frequencies will be lost. Consequently, a simpler method, based only on signal energy, will be required to supplement the existing technique in such cases; such methods are in widespread use in audio analysis [37, 38]. Similarly, in image segmentation the ambiguities which result from reliance on boundary curves could be greatly reduced by using some information from region interiors. Although the inference methods used were based on likelihood maximisation, a number of significant issues in terms of modelling and measurement have not received adequate attention yet. These improvements are the subject of continued study. One of the attractions of using a representation like the MFT is precisely that it allows arbitrary variations of the signal model, without compromising its localisation properties.

Acknowledgement - This work was supported in part by the U.K. SERC and in part by Solid State Logic Ltd of Begbroke, Oxford.

5 Appendix : Proofs of Results in Section 2

Proof of Lemma. For convenience, define $\mathcal{G}_0 = \mathcal{G}$ and $\mathcal{G}_N(\xi_1, \xi_2) = \mathcal{G}_N \cap \mathcal{G}(\xi_1, \xi_2)$ for each finite interval $[\xi_1, \xi_2)$. Suppose the lemma is true for $M < N$ and let $[\xi_N, \xi'_N)$ be the support of g_N. Then there exists for any $f \in L^2(\xi_N, \xi'_N)$ an expansion of the form $f = \sum_{n \geq N} \alpha_n g_n$. Now put $\xi''_N = (\xi_N + \xi'_N)/2$ and write f as $f = f_1 + f_2$, where $f_1 \in L^2(\xi_N, \xi''_N)$ and $f_2 \in L^2(\xi''_N, \xi'_N)$. Then

$$||f||^2 = ||f_1||^2 + ||f_2||^2 \tag{78}$$

and there is an expansion of $f_i, i = 1, 2$ of the form

$$f_i = \sum_{n \geq N} \alpha_n^i g_n \quad i = 1, 2 \tag{79}$$

where $\alpha_n^i = 0, i = 1, 2$, if $g_n \notin \mathcal{G}_{N-1}(\xi_N, \xi_N'') \cup \mathcal{G}_{N-1}(\xi_N'', \xi_N')$. But g_N is non-zero on both these intervals and so $g_N \notin \mathcal{G}_{N-1}(\xi_N, \xi_N'') \cup \mathcal{G}_{N-1}(\xi_N'', \xi_N')$. Hence $\alpha_N^i = 0, i = 1, 2$. Furthermore, since $\mathcal{G}_{N-1}(\xi_N, \xi_N'') \cap \mathcal{G}_{N-1}(\xi_N'', \xi_N') = \emptyset$, $\alpha_n^1 \alpha_n^2 = 0, n \geq N$. As $\mathcal{G}_{N-1}$ is locally bounded with bound A, then from (79)

$$\sum_{n \geq N} |\alpha_n^i|^2 \leq A||f_i||^2 \quad i = 1, 2 \tag{80}$$

Hence, from (78) and $\alpha_n^1 \alpha_n^2 = 0$,

$$\sum_{n \geq N} |\alpha_n^1 + \alpha_n^2|^2 \leq A||f||^2 \tag{81}$$

Thus if $\mathcal{G}_M, M < N$, is locally bounded with bound A, then so is $\mathcal{G}_N$. Since $\mathcal{G}_0$ is so bounded, it follows at once that $\mathcal{G}_N$ is for $N > 0$.

Proof of Theorem. If $\mathcal{G} = \{g_n, n \in Z_+\}$ is a frame, then by Lemma 7.5 of [61], there is for each $f \in L^2(R)$ a unique scalar sequence $\{\alpha_n, n \in Z_+\}$ such that

$$f = \sum_n \alpha_n g_n \tag{82}$$

and, if $\{\beta_n, n \in Z_+\}$ is any other sequence for which $f = \sum_n \beta_n g_n$, then

$$\sum_n |\beta_n|^2 = \sum_n |\alpha_n|^2 + \sum_n |\alpha_n - \beta_n|^2 \tag{83}$$

Now let $\mathcal{G}_N$ be as above the subset of $\mathcal{G}$ from which the first N vectors have been removed. Then by the lemma and Lemma 7.6 of [61], it follows that if $\mathcal{G}$ is a frame, then so is $\mathcal{G}_N, N > 0$. Thus for each N there exists a sequence $\{\alpha_n^N, n \in Z_+\}$, such that $\alpha_n^N = 0, n \leq N$, satisfying (82) above. Now as $\mathcal{G}$ is locally bounded, for any $\epsilon > 0$, there is an $N \geq 0$ for which

$$\sum_{n > N} |\alpha_n^0|^2 < \epsilon \sum_n |\alpha_n^0|^2 \tag{84}$$

But, from (83),

$$\sum_n |\alpha_n^N|^2 = \sum_n |\alpha_n^0|^2 + \sum_{n \leq N} |\alpha_n^0|^2 + \sum_{n > N} |\alpha_n^0 - \alpha_n^N|^2 \tag{85}$$

and clearly

$$\sum_{n > N} |\alpha_n^0 - \alpha_n^N|^2 \geq (1 - \delta)^2 \sum_n |\alpha_n^N|^2 \tag{86}$$

where

$$\delta^2 = \sum_{n>N} |\alpha_n^0|^2 / \sum_n |\alpha_n^N|^2 < \epsilon \sum_n |\alpha_n^0|^2 / \sum_n |\alpha_n^N|^2 \tag{87}$$

Then using (84), (87) and (86) in (85) and using the result of the lemma to show that each $\mathcal{G}_N$ has bound A, it follows that for any $\epsilon > 0$,

$$\sum_n |\alpha_n^0|^2 < \epsilon A ||f||^2 \tag{88}$$

implying that $\alpha_n^0 = 0, n = 0, 1, 2..$, which cannot be true. Hence no such unique sequence exists : $\mathcal{G}$ is not a frame.

Proof of Proposition. In proving this result, there are two cases to consider : rational and irrational scale constant s. Consider first the case of s irrational. It is required to show that the N-consistency ρ_N, defined by

$$\rho_N(s) = \frac{\sum_{j=-N}^{N} \sum_k \mu(I_{j0}^s \cap I_{k1}^s)}{\sum_{j=-N}^{N} \sum_k \mu(I_{j0}^s \cup I_{k1}^s) \nu(I_{j0}^s \cap I_{k1}^s)} \tag{89}$$

possesses a limit as $N \to \infty$. Since the intervals $I_{kj}, k \in Z$ are disjoint and cover R, for $j = 0$ or $j = 1$, the numerator in (89) can be rewritten as

$$\sum_{j=-N}^{N} \sum_k \mu(I_{j0}^s \cap I_{k1}^s) = \sum_{j=-N}^{N} \mu(I_{j0}^s) = (2N + 1) \tag{90}$$

In the interval $[-N, N)$, there are $\lfloor(2N+1)s\rfloor + 2$ intervals of the form $[ls^{-1}, (l+1)s^{-1})$, of which exactly $2N$ bracket the integers $k = \pm 1, \pm 2, .., \pm N$ because s is irrational. Each such interval bracketing an integer gives rise to a contribution $(2 + s^{-1})$ to the sum in the denumerator since it overlaps the two intervals $[k - 1, k)$ and $[k, k + 1)$. The intervals overlapping the ends of the range, on the other hand, give an average contribution of 1/2 of that amount, since only one of the intersections is nonempty. Each of the remaining intervals gives rise to a contribution of 1, since it is contained within some interval of the form $[k, k+1)$. The sum in the denumerator can therefore be written as

$$\sum_{j=-N}^{N} \sum_k \mu(I_{j0}^s \cup I_{k1}^s) \nu(I_{j0}^s \cap I_{k1}^s) = \lfloor(2N + 1)s\rfloor + 2(N - 1)(1 + s^{-1}) + s^{-1} \tag{91}$$

Substituting from (91) and (90) into (89) and taking the limit as $N \to \infty$ gives

$$\rho(s) = \lim_{N \to \infty} \rho_N(s) = \frac{s}{1 + s + s^2} \tag{92}$$

Since $s > 1$, $\rho(s) < 1/2$ for any irrational s.

When $s = n/m$, where $m < n$ and m and n are mutually prime, there are $\lfloor (2N+1)n/m \rfloor + 2\eta(N/m)$ intervals of the form $[lm/n, (l+1)m/n)$ in $[-N, N)$, where $\eta(x) = 0$ if $x = \lfloor x \rfloor$ and $\eta(x) = 1$ otherwise. Of these, $2\lfloor N/m \rfloor + 1$ have an end point which is an integer. Following similar reasoning to that above, the denumerator of (89) becomes

$$\sum_{j=-N}^{N} \sum_{k} \mu(I^s_{j0} \cup I^s_{k1}) \nu(I^s_{j0} \cap I^s_{k1}) = \lfloor (2N+1)n/m \rfloor + 2\eta(N/m) + 2(N - \lceil N/m \rceil)(1 + m/n) + \eta(N/m)m/n \tag{93}$$

Substituting from (93) and (90) into (89) and taking the limit gives

$$\rho(n/m) = \frac{mn}{n^2 + m^2 + mn - m - n} \tag{94}$$

If $\rho(n/m) > 1/2$ then from (94)

$$2mn > n^2 + m^2 + mn - m - n \tag{95}$$

or

$$m^2 - m(n-1) + (n-1)^2/4 < (n-1)^2/4 + n - n^2 \tag{96}$$

giving

$$3n^2 - 2n < 1 \tag{97}$$

implying that $n < 1$, which is impossible. Thus $\rho(n/m) \leq 1/2$. If $\rho(n/m) = 1/2$, then the inequality in (95) is replaced by equality, giving

$$m^2 - m(n+1) + n(n-1) = 0 \tag{98}$$

and solving for m gives

$$m = (n+1)/2 \pm \sqrt{1 + 6n - 3n^2}/2 \tag{99}$$

which has a meaningful real solution only when $n = 2$, giving $\rho(2) = 1/2$, which completes the proof of the proposition.

References

[1] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis", *IEEE Trans. Inform. Th.*, vol. IT-36, pp. 961-1005, 1990.

[2] P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code ", *IEEE Trans. Comun.*, vol. COM-31, pp. 532-40, 1983.

[3] E.H. Adelson, E. Simoncelli, and R. Hingorani, "Orthogonal pyramid transforms for image coding", *Proc. SPIE Colloq. on Vis. Commun. and IP*, Cambridge, MA, 1987.

[4] A. Grossman and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape", *SIAM J. Math. Anal.*, 15, pp. 723-736, 1984.

[5] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless non-orthogonal expansions", *J. Math. Phys.*, 27, pp. 1271-83, 1986.

[6] S. Mallat, "A theory for multiresolution signal decomposition : The wavelet representation", *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. PAMI-11, pp. 674-93, 1989.

[7] R. Wilson and M. Spann, "Finite prolate spheroidal sequences and their applications, pt II", *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-10, pp. 193-203, 1988.

[8] M. Porat and Y.Y. Zeevi, "The generalized Gabor scheme of image representation in biological and machine vision", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-10, pp. 452-68, 1988.

[9] D. Slepian, "On bandwidth", *Proc. IEEE*, vol. 64, pp. 292-300, 1976.

[10] M.R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis", *IEEE Trans. Acous. Speech Sig. Proc.*, vol. ASSP-28, pp. 55-69, 1980.

[11] D.L.Jones and T.W.Parks, "A high resolution data-adaptive time-frequency representation", *IEEE Trans. Acous. Speech Sig. Proc.*, vol. ASSP-38, pp. 2127-2135, 1990.

[12] R. Wilson and H.E. Knutsson, "Uncertainty and inference in the visual system", *IEEE Trans. Sys. Man Cybern.*, vol. SMC-18, pp. 305-12, 1988.

[13] D. Slepian, "Prolate spheroidal wavefunctions, Fourier analysis, and uncertainty - V : the discrete case",*Bell Syst. techn. J.*, 57, pp. 1371-1429, 1978.

[14] R. Wilson and A.D. Calway, "A general multiresolution signal descriptor and its application to image analysis", *Proc. EUSIPCO-88*, pp. 663-666, Grenoble, 1988.

[15] A.D. Calway, *The Multiresolution Fourier Transform: A General Purpose Tool for Image Analysis*, Ph.D. Thesis, Warwick Univ., 1989.

[16] A.D. Calway and R. Wilson, "Curve extraction in images using the multiresolution Fourier transform", *Proc. IEEE ICASSP-90*, pp. 2129-2132, Albuquerque, 1990.

[17] E.R.S. Pearson,*The Multiresolution Fourier Transform and its Application to the Analysis of Polyphonic Music*, Ph.D. Thesis, Warwick Univ., 1991.

[18] R. Wilson, A.D. Calway and E.R.S. Pearson, "A generalized wavelet transform for Fourier analysis : the Multiresolution Fourier Transform and its application to image and audio signal analysis",To be published *IEEE Trans. Inform. Th.*.

[19] A.R. Davies," Curve and corner extraction using the MFT",*Warwick Univ. Computer Science Dept. Rpt. no.* ,Warwick,1991.

[20] A. Witkin, "Scale-space filtering", *Proc. IEEE ICASSP-84*, San Diego, 1984.

[21] S. Mallat and S. Zhong, "Complete signal representation with multiscale edges", *NYU Comput. Sci. Rept. 483*, 1989.

[22] F.J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform", *Proc. IEEE*, vol. 66, pp. 51-83, 1978.

[23] N.I. Akhiezer and I.M. Glazman, transl. M. Nestell, *Theory of Linear Operators in Hilbert Space vol. II*, New York, Fred. Ungar,1963.

[24] D. Slepian and H.O. Pollak,"Prolate spheroidal wavefunctions, Fourier analysis and uncertainty I",*Bell Syst. techn. J.*,vol. 40,pp. 43-64, 1961.

[25] A. Papoulis, *Signal Analysis*, New York: McGraw-Hill, 1977.

[26] D.J. Thomson, "Spectrum estimation and harmonic analysis", *Proc. IEEE*, vol. 70, pp. 1055-1096, 1981.

[27] I. Daubechies, "Orthonormal bases of compactly supported wavelets", *Commun. Pure Appl. Math.*, vol. 41, pp. 909-96, 1988.

[28] J.P. Elliott and P.G. Dawber, *Symmetry in Physics vol. 1*, London: Macmillan,1979.

[29] L.R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Englewood-Cliffs: Prentice-Hall, 1975.

[30] R. Wilson, M. Spann, *Image Segmentation and Uncertainty*, Letchworth, Res. Studies Pr., 1988.

[31] D. Marr, *Vision*, San Francisco: Freeman, 1982.

[32] H.E. Knutsson, R. Wilson and G.H. Granlund, " Anisotropic nonstationary image estimation and its applications", *IEEE Trans. Commun.*,vol. COM-31, pp. 388-406,1983.

[33] G.H. Granlund," In search of a general picture processing operator", *Comput. Graph. Image Proc.*,vol. 8,pp. 155-73,1978.

[34] S. Geman and D. Geman," Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images",*IEEE Trans. Patt. Anal. and Machine Intell.*,6,721-741,1984.

[35] A. Rosenfeld, Ed., *Multiresolution Image Processing and Analysis.* Berlin, Germany: Springer, 1984.

[36] J.A. Moorer, *On the Segmentation and Analysis of Continuous Musical Sound by Digital Computer*, Ph.D. Thesis, Stanford Univ., 1975.

[37] C. Watson, *The Computer Analysis of Polyphonic Music*, Ph.D. Thesis, Sydney Univ., 1986.

[38] X. Serra, *A System For Sound Analysis/Transformation/ Synthesis Based on a Deterministic plus Stochastic Decomposition*, Ph.D. Thesis, Stanford Univ., 1989.

[39] C. Chafe, D. Jaffe, K. Kashima, B. Mont-Reynaud, and J.O.Smith, "Techniques for note identification in polyphonic music", *Proc. Int'l. Conf. Comput. Music*, 1985.

[40] L.R. Bahl, F. Jelinek, and R.L. Mercer, "A maximum likelihood approach to continuous speech recognition", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 179-90, 1983.

[41] T.W. Anderson, *The Statistical Analysis of Time Series*, New York: Wiley, 1971.

[42] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 2nd Ed., New York: McGraw-Hill, 1984.

[43] A.K. Jain,*Fundamentals of Digital Image Processing*,Englewood-Cliffs: Prentice-Hall,1989.

[44] R.C. Gonzalez and P. Wintz,*Digital Image Processing*,2nd Ed.,Reading, Mass. : Addison-Wesley,1987.

[45] T.R. Reed and H. Wechsler,"Segmentation of textured images and gestalt organization using spatial/spatial frequency representations",*IEEE Trans. Pattern Anal. Machine Intell.*,vol. PAMI-12,pp. 1-12,1990.

[46] J.G.Daugman, "Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 1169-1179, 1988.

[47] J.J. Koenderink, "The structure of images", *Biol. Cybern.*, 50, pp. 363-370, 1984.

[48] U. Montanari, "On the optimal detection of curves in noisy pictures", *Commun. ACM*, 14, pp. 335-345, 1971.

[49] G.P. Ashkar and J.W. Modestino," The contour extraction problem with biomedical applications,*Comput. Graph. Image Proc.*,vol. 7,pp. 331-55, 1978.

[50] M. Kass, A. Witkin, and D. Terzopoulos, "SNAKES: active contour models", *Int. J. Computer Vision*, vol. 1, pp. 321-332, 1988.

[51] S.W. Zucker, C. David, A. Dobbins, and L. Iverson, "The organization of curve detection: Coarse tangent fields and fine spline coverings", *Proc. Int. Conf. Computer Vision*, pp. 568-577, 1988.

[52] D.G. Lowe,"Organization of smooth curves at multiple scales",*Proc. Int'l. Conf. Comput. Vision*,pp.558-67,1988.

[53] S. Connelly and A. Rosenfeld, "A pyramid algorithm for fast curve extraction", *Comput. Vision, Graph. Image Proc.*, vol. 49, pp. 332-345, 1990.

[54] P. Grattoni, F. Pollastri and A. Premmoli," A contour detection algorithm based on the minimum radial inertia (MRI) criterion",*Comput. Vision, Graph. Image Proc.*,Vol. 43,pp. 23-36,1988.

[55] J. Princen, J. Illingworth, and J. Kittler, "A hierarchical approach to line extraction based on the Hough transform", *Comput. Vision, Graph. Image Proc.*, vol. 52, pp. 57-77, 1990.

[56] H. Knutsson, "Representing local structure using tensors", *Proc. 6th Scan. Conf. on Image Analysis*, Oulu, 1989.

[57] T.W. Anderson,*An Introduction to Multivariate Statistical Analysis*, New York, Wiley, 1958.

[58] P-E. Danielsson,"Rotation invariant linear operators with directional response",*Proc. 5th Int'l Conf. Patt. Recogn.*,pp. 1171-76,Miami,1980.

[59] R. Lenz,"Optimal filters for the detection of linear patterns in 2-d and higher dimensional spaces",*Patt. Recogn.*,vol. 20,pp. 163-72,1987.

[60] J.M. Grey, J.W. Gordon, "Perception of Spectral Modifications on Orchestral Tones", *Comput. Music J.*, vol. 2, pp. 24-31, 1978.

[61] R.M. Young,*An Introduction to Nonharmonic Fourier Series*, New York: Academic Pr., 1980.

[62] A.R. Davies, R. Wilson,"Curve and Corner Extraction Using the Multiresolution Fourier Transform",*Proc. IEE Conf. on Image Proc. and Applications*, Maastricht, Holland, 1992.