

Ye Wang and Miikka Vilermo
Nokia Research Center
FIN-33720 Tampere, Finland

Leonid Yaroslavsky
Tel Aviv University
Ramat Aviv 69978, Israel

**Presented at
the 109th Convention
2000 September 22-25
Los Angeles, California, USA**



AES

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd St., New York, New York 10165-2520, USA.

All rights reserved. Reproduction of this preprint, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

AN AUDIO ENGINEERING SOCIETY PREPRINT

ENERGY COMPACTION PROPERTY OF THE MDCT IN COMPARISON WITH OTHER TRANSFORMS

Ye Wang¹, Miikka Vilermo¹, Leonid Yaroslavsky²

¹Nokia Research Center,
Speech and Audio Systems Laboratory,
FIN-33720 Tampere,
Finland

²Department of Interdisciplinary Studies,
Tel Aviv University,
Ramat Aviv 69978,
Israel

e-mail: ye.wang@nokia.com, miikka.vilermo@nokia.com, yaro@eng.tau.ac.il

Abstract - This paper focuses on the energy compaction properties of five different transforms: DFT, DCT, SDFT((N+1)/2, 1/2), MDCT and DST. Energy compaction properties of these transforms are compared experimentally. In addition to sinusoidal signals, sixteen classical and pop music pieces are used for the experiments. The influence of different window sizes (256, 512 and 1024 samples) and different window shapes (rectangular and sine) are investigated. The results of the experiments are presented and analyzed.

I. INTRODUCTION

Signal Fourier spectrum analysis is one of the major tools of signal processing. For real-life continuous signals such as audio signals and images, it is associated with signal integral Fourier transformation. In digital signal processing, integral Fourier transformation is approximated by Discrete Fourier Transforms implemented via Fast Fourier Transform algorithms. On the other hand, it has been found that in image and audio coding, restoration and similar applications other transforms such as Discrete Cosine Transform (DCT), Discrete Sine Transform (DST), and Modified Discrete Cosine Transform (MDCT) may be more suitable than DFT [1][2]. However, it is often necessary to establish interrelations between DFT signal spectra and those of DCT, MDCT, and DST to evaluate their applicability for signal Fourier analysis. Based on these interrelations, the energy compaction properties of these transforms are investigated in this paper. This work is an extension of our previous paper [3].

In most state-of-the-art audio encoders, MDCT [4] is used to compress signals in the frequency domain. In this context, it is necessary to examine how well the transform approximates the Fourier spectrum and why the MDCT exhibits an energy compaction property. Although MDCT fails in some special situations [5] for spectral analysis, it is commonly used in audio coding applications.

Transform energy compaction capability means the capability of the transform to redistribute signal energy into a small number of transform coefficients. It can be characterized by the fraction of the total number of signal transform coefficients that carry a certain (substantial) percentage of the signal energy. The lower this fraction is for a given energy percentage, the better the transform energy compaction capability is.

There are different approaches to studying the energy compaction property of different transforms, because the spectral discretization interval of the transforms may be different. In the case of stationary signals, the conventional solution is to use different time-domain window sizes so that the spectral discretization intervals of different transforms remain the same. However, if the signal is nonstationary it may be more reasonable to use the same time-domain window size for all transforms, because time domain windows of different sizes may contain significantly different frequency components. In addition there are certain constraints on the window size in different applications. In our approach, we employ interpolation and normalization to align all transform spectra in the same coordinate for a fair comparison.

Purely analytical evaluation of the transform energy compaction capability is problematic since it is only feasible for limited mathematical models of signals. Another option is to evaluate the energy compaction property experimentally with a large number of test signals. In this paper, we first study the energy compaction property of different transforms by comparing the spectral resolution of individual sinusoids. Then we present the results of such an evaluation for a set of 8 pieces of classical music and 8 pieces of pop music. For these signals, the energy compaction capability of transforms is investigated with different window sizes (256, 512 and 1024 samples) and with different window functions (rectangular and sine) over 60*44100 samples. The sampling frequency was 44.1 kHz. The results are illustrated in frequency coordinates normalized to [0-1] by the Nyquist frequency.

II. INTERRELATION BETWEEN INTEGRAL FOURIER TRANSFORM, DFT, DCT, MDCT, DST

Discrete representation of signal integral transforms parallels that of signals. For a signal $a(x)$ and its Fourier spectrum $\alpha(f)$ represented in a discrete form by means of sequences of their samples $\{a_k\}$ and $\{\alpha_r\}$ taken at sets of equidistant points $\{(k+u)\Delta x\}$ and $\{(r+v)\Delta f\}$, $k = \dots, -2, -1, 0, 1, 2, \dots$; $r = \dots, -2, -1, 0, 1, 2, \dots$ such that

$$a(x) = \sum_k a_k \varphi_x(x - (k+u)\Delta x), \quad (1)$$

$$\alpha(f) = \sum_r \alpha_r \varphi_f(f - (r+v)\Delta f), \quad (2)$$

where Δx and Δf are discretization intervals and u and v are shifts (in a fraction of the corresponding discretization interval) of sample positions from the origin of the corresponding coordinates, discrete representation of the Fourier integral

$$\alpha(f) = \int_{-\infty}^{\infty} a(x) \exp(i2\pi fx) dx \quad (3)$$

takes the form of ‘‘Shifted Discrete Fourier Transforms’’ (SDFT) [2]:

$$\alpha_r = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a_k \exp\left(i2\pi \frac{kv}{N}\right) \exp\left(i2\pi \frac{(k+u)r}{N}\right) \quad (4)$$

the most wide known special case of which (for zero shifts u and v) is DFT:

$$\alpha_r = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a_k \exp\left(i2\pi \frac{kr}{N}\right) \quad (5)$$

Other well-known transforms in digital signal processing such as DCT, MDCT and DST are also special cases of SDFT. From derivation it can be seen that signal spectra obtained by DCT, MDCT, DST are identical to Shifted Discrete Fourier Transform spectra of certain permutation modifications of the original signal:

$$\text{DCT: } \alpha_r = \sum_{k=0}^{N-1} a_k \cos\left[\pi \frac{\left(k + \frac{1}{2}\right)r}{N}\right] = \frac{1}{2} \sum_{k=0}^{2N-1} \tilde{a}_k \exp\left[i2\pi \frac{\left(k + \frac{1}{2}\right)r}{2N}\right], \quad (6)$$

$$\text{where } \tilde{a}_k = \begin{cases} a_k, & k = 0, \dots, N-1 \\ a_{2N-1-k}, & k = N, \dots, 2N-1 \end{cases}$$

$$\text{MDCT: } \alpha_r = \sum_{k=0}^{N-1} a_k \cos \left[2\pi \frac{\left(k + \frac{N+2}{4}\right) \left(r + \frac{1}{2}\right)}{N} \right] = \frac{1}{2} \sum_{k=0}^{N-1} \tilde{a}_k \exp \left[i2\pi \frac{\left(k + \frac{N+2}{4}\right) \left(r + \frac{1}{2}\right)}{N} \right], \quad (7)$$

where $\tilde{a}_k = \begin{cases} a_k - a_{N/2-1-k}, & k = 0, \dots, N/2-1 \\ a_k + a_{3N/2-1-k}, & k = N/2, \dots, N-1 \end{cases}$, and N is even;

$$\text{DST: } \alpha_r = \sum_{k=0}^{N-1} a_k \sin \left[\pi \frac{(k+1)(r+1)}{N+1} \right] = \frac{1}{2} \sum_{k=0}^{2N} \tilde{a}_k \exp \left[i2\pi \frac{(k+1)(r+1)}{2(N+1)} \right], \quad (8)$$

Where $\tilde{a}_k = \begin{cases} a_k, & k = 0, \dots, N-1 \\ 0, & k = N, 2N+1 \\ -a_{2N-k}, & k = N+1, \dots, 2N \end{cases}$

These relationships mnemonically illustrated in Fig. 1 lucidly explain the interrelations between the above trigonometric bases and their similarity and dissimilarity.

III. COMPARISON OF TRANSFORM SPECTRAL RESOLUTION POWER AND ENERGY COMPACTION CAPABILITY

In this section, we compare the above trigonometric bases in terms of their energy compaction capability and their resolution power in Fourier spectrum analysis. This property is relevant for many applications, such as signal compression/coding.

DFT, DCT, DST and MDCT all have different spectral discretization intervals. For an N-samples long real sequence the independent DFT bins represent frequencies $2k/N, k = 0, \dots, \lfloor N/2 \rfloor$. The frequency ordinate is normalized to the Nyquist frequency for simplicity. DCT bins represent frequencies $k/N, k = 0, \dots, N-1$, DST bins represent frequencies $(k+1)/(N+1), k = 0, \dots, N-1$, and MDCT bins represent frequencies $(2k+1)/N, k = 0, \dots, N/2-1$. For MDCT we assume N to be even since MDCT is a Lapped Orthogonal Transform (LOT). The discretization interval Δf of DFT is the basis of all comparisons described in this paper. Note that the discretization intervals of DFT and MDCT are twice as long as that of DCT and DST.

Transform Resolution Power with Sinusoids

The transform resolution power in signal spectral estimation characterizes the sharpness of spectral peaks of sinusoidal signals. It can be evaluated numerically as the width, in proportion to the discretization interval, of the spectral peak within which a given (substantial) percentage of the energy of a sinusoidal signal is contained. From sampling theory it follows that the width of the spectral peaks in the signal discrete spectrum is, in general, proportional to the discretization interval in the frequency domain. However, the proportionality is different for different discrete trigonometric transforms discussed in previous section.

Evaluation of transform spectral resolution power requires testing the spectral peak width of sinusoidal signals having arbitrary frequencies within the frequency range defined by the signal discretization rate. Although the evaluation can be carried out analytically in principle, the same results can be obtained by numerical simulation of the transforms. As initial numerical simulations, sine test signals with frequencies uniformly distributed within the corresponding frequency discretization interval were selected and the results of spectrum estimation for each central frequency were published in [3]. Those results were averaged in such a way that the spectra within a discretization interval were added and the resulting spectrum was used to measure the resolution power. 100 realizations were used for each spectral discretization interval.

In this paper, we have chosen an improved method to estimate the resolution power. Instead of adding the 100 spectra together, we measure the individual spectral width of the 100 realizations, and then average them within each frequency discretization interval. In addition, we have tested cosine signals and cosine signals with random phase shifts. Some new results are reported in this paper.

The principle of our method is illustrated in Fig. 2 and Fig. 3 using only 18 time domain samples for clarification. The dashed lines in Fig. 2 represent the actual spectral lines of a sinusoid $\sin(2\pi ft)$ whose frequency changes within one frequency discretization interval Δf . DFT and DCT spectra (black and white respectively in Fig. 2) of $f_r - \Delta f/2$, $f_r - \Delta f/4$, f_r , $f_r + \Delta f/4$, $f_r + \Delta f/2$ are illustrated in (a)-(e), where f_r corresponds to one DFT sampling point. In order to have the same spectral discretization interval, DFT is interpolated by a factor of 2 using the lowpass interpolation algorithm described in [6]. Obviously these two spectra are different representations of the same sinusoid. For more precision, both DFT and DCT spectra are further interpolated by a factor of 5 in Fig. 3. Then we set the energy threshold at 50% and measure the normalized width of each spectrum and then take the averaged value within each frequency discretization interval Δf .

By increasing the realization from 5 to 25, the development of the DFT spectral shape of a cosine signal with a frequency changing from $f_r - \Delta f/2$ to $f_r + \Delta f/2$ is illustrated in Fig. 4. Similarly, the development of the DCT spectral shapes of the same cosine signal is illustrated in Fig. 5-7, MDCT spectral shapes in Fig. 9-11. Interestingly, the pattern of the DFT spectral shapes within each Δf remains the same in all frequency regions, while the patterns of DCT and MDCT spectral shapes within each Δf change with frequency.

The averaged frequency resolutions of the transforms are illustrated in Fig. 12-16 using sine, cosine and cosine with a random phase shift respectively. On average in the whole frequency range [0-1] with rectangular window as in Fig. 12, 14, 15, the frequency resolutions are $0.6635\Delta f$ for DFT, $0.7171\Delta f$ for MDCT, $0.525\Delta f$ for DCT and $0.5286\Delta f$ for DST. However, applying sine window has changed the frequency resolution landscape as illustrated in Fig. 13, 16.

From (6) (8) it can be seen that the phase of a_k has a direct impact on the DCT and DST coefficients. If a_k is in phase with the basis function, the frequency resolution of the transform is optimal. Conversely, a phase shift of $\frac{\pi}{2}$ corresponds to the most sub-optimal frequency resolution. This is verified by the experiments. To explain the frequency-dependent resolution power of DCT, we take the DCT basis function from (6) and change the expression to:

$$\cos\left[\pi\frac{\left(k+\frac{1}{2}\right)r}{N}\right] = \cos\left[\frac{2\pi kr/2}{N} + \frac{\pi r}{2N}\right] \quad (9)$$

The second term in the right hand side bracket is the phase of the basis function $\varphi = \frac{\pi r}{2N}$, $r = 0, \dots, N-1 \Rightarrow$

$$\varphi = 0, \dots, \frac{\pi(N-1)}{2N} \quad (10)$$

This phase shift of the basis functions explains the frequency dependent resolution power of DCT in Fig 12, 14.

Similarly, we take the DST basis function from (8) and change the expression to:

$$\sin\left[\pi\frac{(k+1)(r+1)}{N+1}\right] = \sin\left[\frac{2\pi k(r/2+1/2)}{N} \frac{N}{N+1} + \frac{\pi(r+1)}{N+1}\right] \quad (11)$$

The second term in the right hand side bracket is the phase of the basis function $\varphi = \frac{\pi(r+1)}{N+1}$, $r = 0, \dots, N-1, \Rightarrow$

$$\varphi = \frac{\pi}{N+1}, \dots, \frac{\pi N}{N+1} \quad (12)$$

If N is big (512 in our experiments), the range of the phase is between 0 and π . This explains the DST frequency dependent resolution power in Fig. 12, 14. Fig 13 illustrates the effect of the window function.

The basis function of DFT is an exponential function, which can be split into sine and cosine basis functions. The sine and cosine basis functions complement each other in the resolution power. This explains why DFT frequency resolution is not sensitive to the phase of the signal as illustrated in Fig. 12-16.

From Fig. 2 to 14 the sinusoids used are all of fixed phase. Fig. 15 and 16 show the frequency resolution of sinusoids having random phase shifts. Because of the special properties of the MDCT [5], it is difficult to explain its frequency dependent resolution analytically, and therefore it is omitted in this paper.

Transform Energy Compaction with Real-life Audio Signals

This section discusses the transform energy compaction property with real-life audio signals. 16 pieces of pop and classic music signals were used in the experiment. Fig. 17 shows the energy compaction property with 8 pieces of pop music signals. The window size is 256-1024 samples for all transforms. Rectangular and sine windows are used. The length of the audio signals are 60*44100 samples and the sampling frequency is 44.1 kHz.

Fig. 18 shows a zoomed version of the comparison, but with a window length of 256 samples. In general, the DCT performs better than other transforms, and the DST performs poorest. This is not very consistent in comparison with the frequency resolution of sinusoidal test signals.

Fig 19 shows the comparison when a sine window is applied. This shows the effect of the window function. As in Fig. 19 the energy compaction property gets more unified with the sine window. Similar comparison with classic music signals is shown in Fig. 20 and 21.

We have also taken the conventional approach to compare the energy compaction property with different time domain windows. The results are shown in Fig 22 - 23. Interestingly, DCT with a window length of 512 performs better than MDCT with a window length of 1024, if rectangular windows are used. However, when keeping the window size unchanged and applying a sine window to both DCT and MDCT or to MDCT only, the energy compaction performance of MDCT is better than DCT. This comparison has been considered to be useful in audio coding applications. Fig. 24 and 25 show the comparison for the case that DCT has a rectangular window length of 512 and MDCT has a sine window length of 1024. However, the different time window may contain significantly different frequency components as noted earlier.

IV. CONCLUSION

All above-mentioned transforms can be used for signal Fourier analysis.

The transforms exhibit different Fourier spectrum analysis resolution power and energy compaction property; the resolution power is not uniform over the entire frequency range for DCT, DST and MDCT using sine and cosine test signals. The averaged resolution power of DFT is uniform within the whole frequency range. On average, over the whole frequency range, DCT and DST have the best resolution power, and MDCT has the poorest resolution power using rectangular window. All these transforms have almost the same resolution power when a sine window is used.

For real-life audio signals, DCT, MDCT and DST exhibit, on average, over large signal sequences, effectively similar energy compaction capabilities. More than 90% energy is concentrated within 10% of the normalized frequency scale for most of the test signals for all transforms concerned. The energy compaction property of different transforms gets more unified with increased window size.

V. ACKNOWLEDGEMENT

Ye Wang wishes to thank Prof. Peter J. Sherman (Iowa State University, USA), Dr. Jilei Tian (Nokia Research Center, Finland), Dr. Bernd Edler (Hannover University, Germany) and Dr. Juergen Herre (FhG-IIS, Germany) for

helpful discussions during ICASSP2000 in Istanbul. The financial support from Nokia Foundation and the Academy of Finland is greatly acknowledged. Leonid Yaroslavsky wishes to thank Tampere International Center for Signal Processing for supporting this research.

REFERENCES

- [1] Malvar, H., "Signal Processing with Lapped Transform", Artech House, Boston, 1991
- [2] Yaroslavsky, L., Eden, M., "Fundamentals of Digital Optics", Birkhauser, Boston, 1996.
- [3] Yaroslavsky, L., Wang, Y., "DFT, DCT, MDCT, DST and Signal Fourier Spectral Analysis", X European Signal Processing Conference (EUSIPCO 2000), Tampere, Finland, September 4-8, 2000.
- [4] Princen, J. P., Bradley, A. B., "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-34, No. 5, October 1986.
- [5] Wang, Y., Yaroslavsky, L., Vilermo, M., Väänänen, M. "Some Peculiar Properties of the MDCT", WCC2000 – 16th IFIP World Computer Congress/ICSP 2000 – 5th International Conference on Signal Processing, August 21 – 25, 2000, Beijing, China.
- [6] Wiley John & Sons, "Programs for Digital Signal Processing", IEEE Press, New York, 1979.

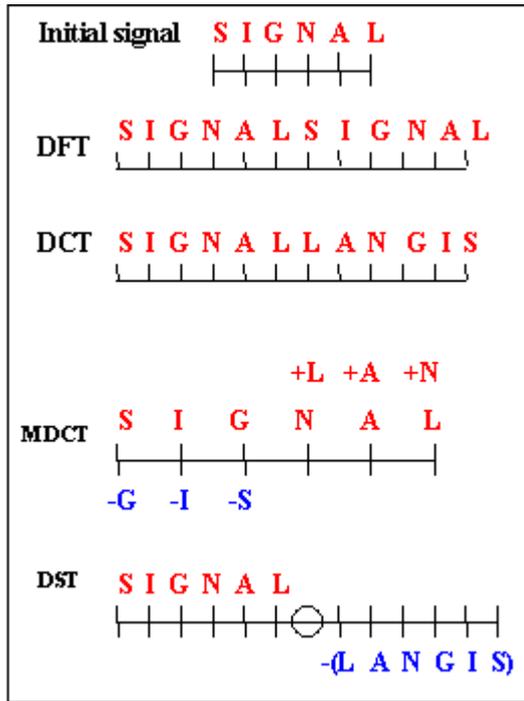


Fig. 1. Signal and its corresponding representations of DFT, DCT, MDCT and DST

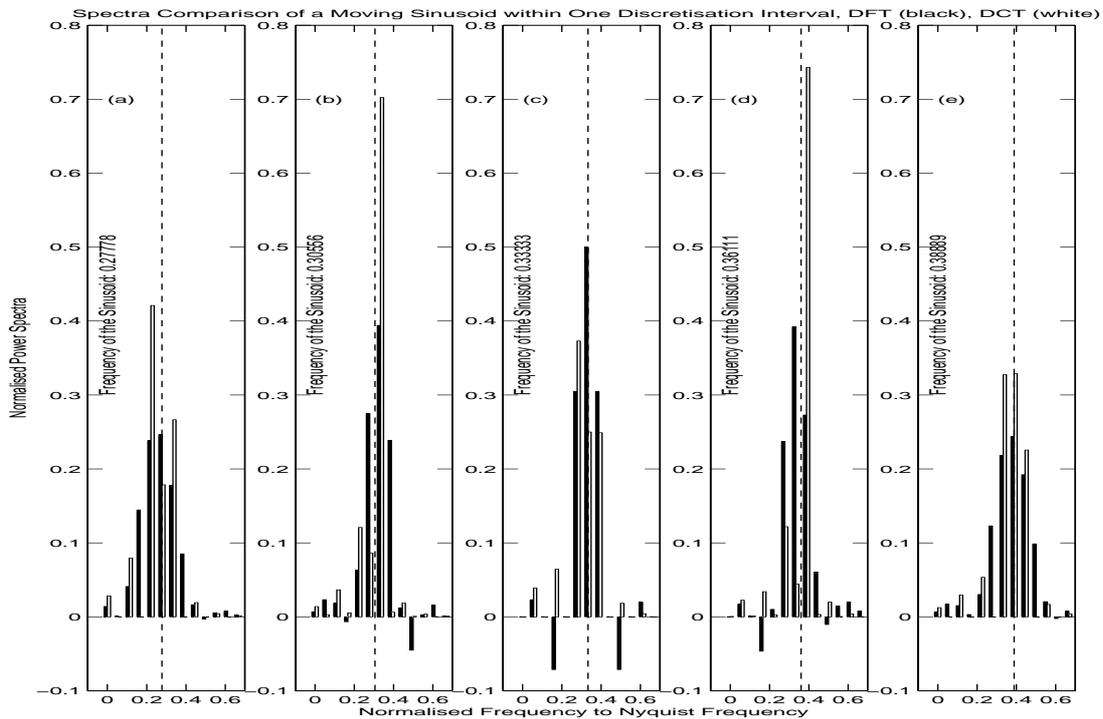


Fig. 2. DFT (black) and DCT (white) power spectra comparison using a sine signal whose frequency (indicated as the dashed lines) changes within one DFT discretisation interval Δf . DFT spectrum is interpolated by factor 2 to have the same spectral discretization interval as DCT. The negative values of DFT power spectra are caused by interpolations.

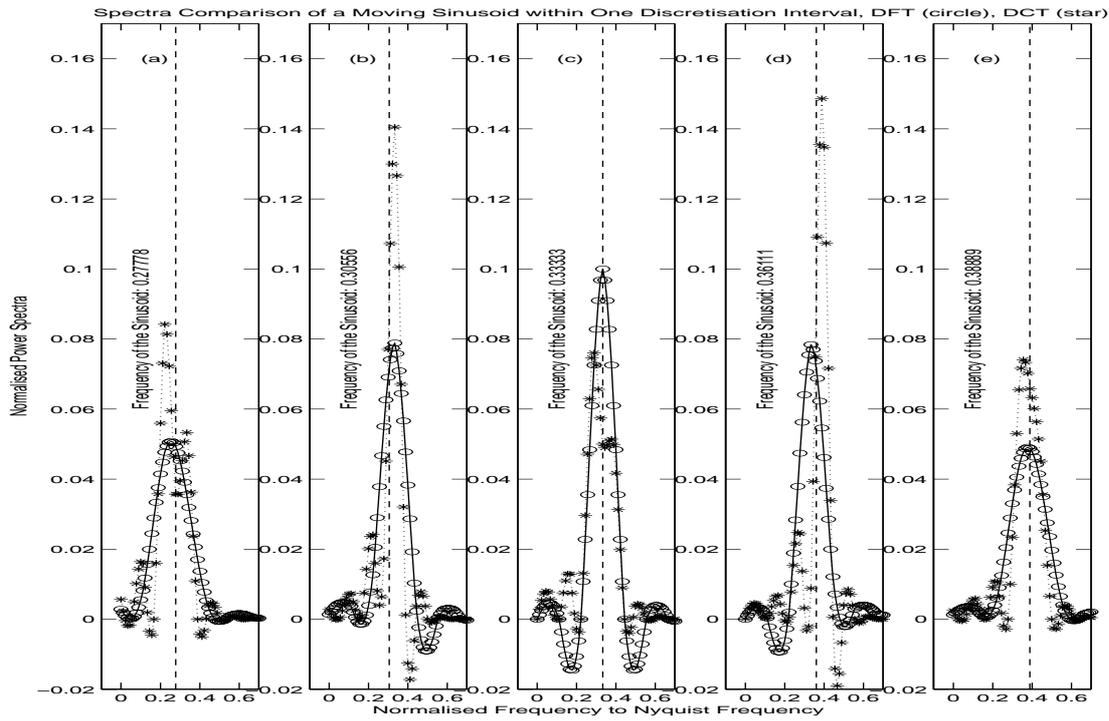


Fig. 3. DFT and DCT power spectra comparison using a sine signal whose frequency (indicated as the dashed lines) changes within one DFT discretisation interval Δf . Both DFT and DCT spectra are further interpolated by factor 5 from Fig.2.

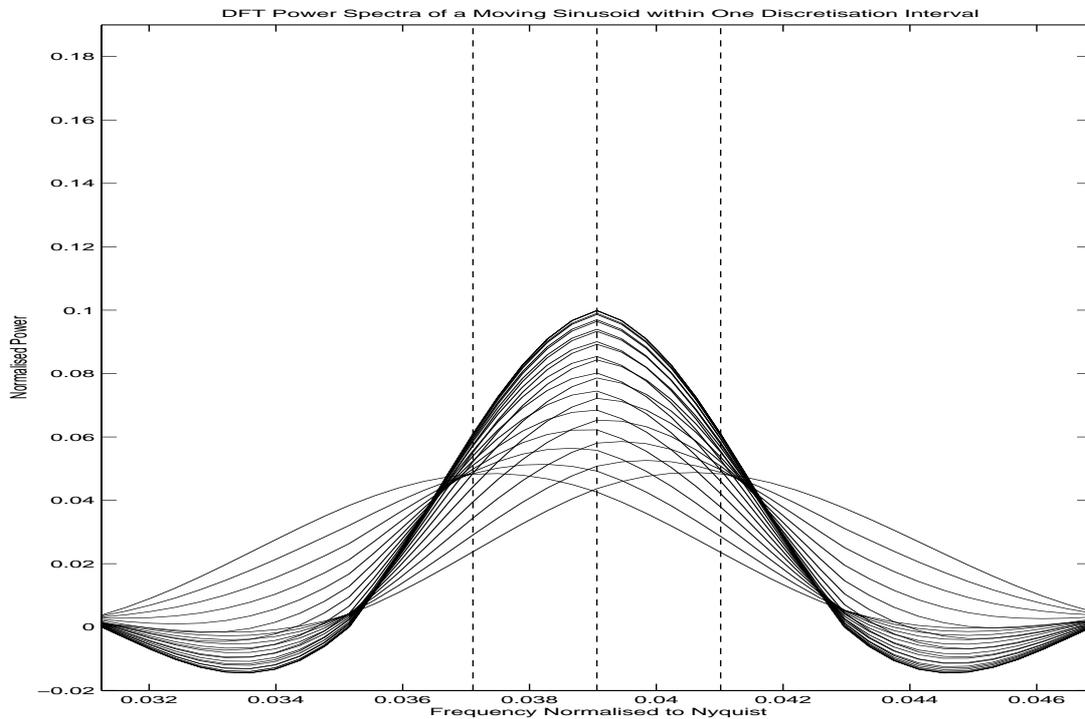


Fig. 4. DFT power spectra of a cosine signal whose frequency changes within one discretisation interval Δf . The 3 dashed lines from left to right represent $f_r - \Delta f/2$, f_r , $f_r + \Delta f/2$.

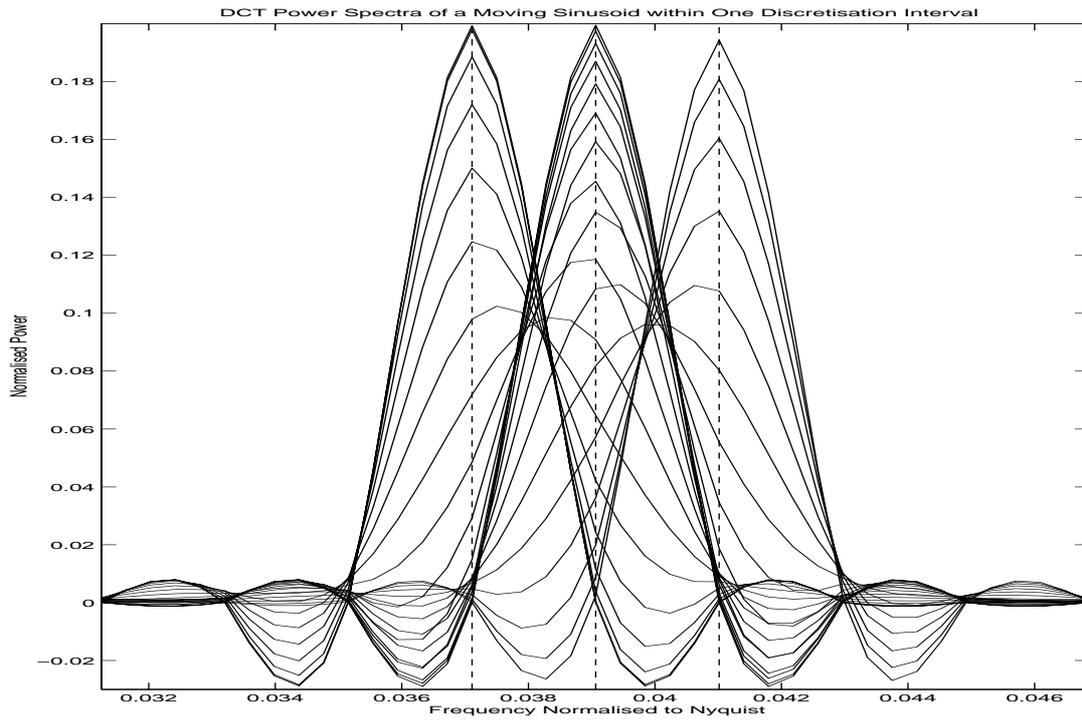


Fig. 5. DCT power spectra of a cosine signal whose frequency changes within Δf in low frequency range. Note that Δf is twice as long as DCT discretization interval. The dashed lines correspond to the DCT spectral sampling points.

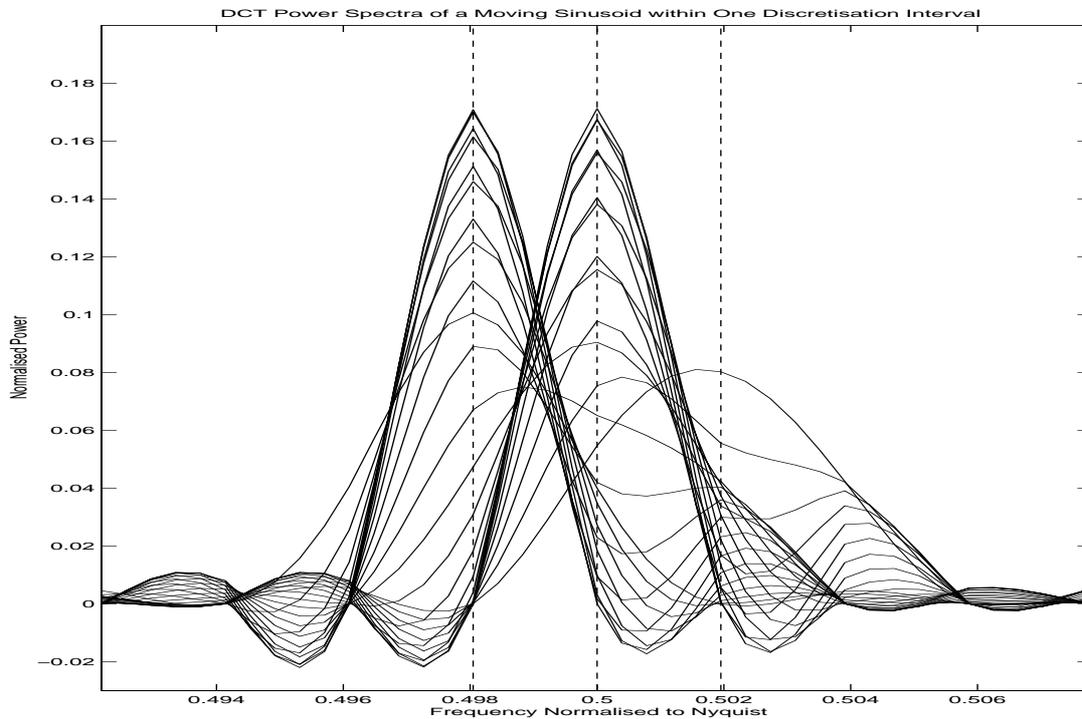


Fig. 6. DCT power spectra of a cosine signal whose frequency (around half of the Nyquist frequency) changes within Δf .

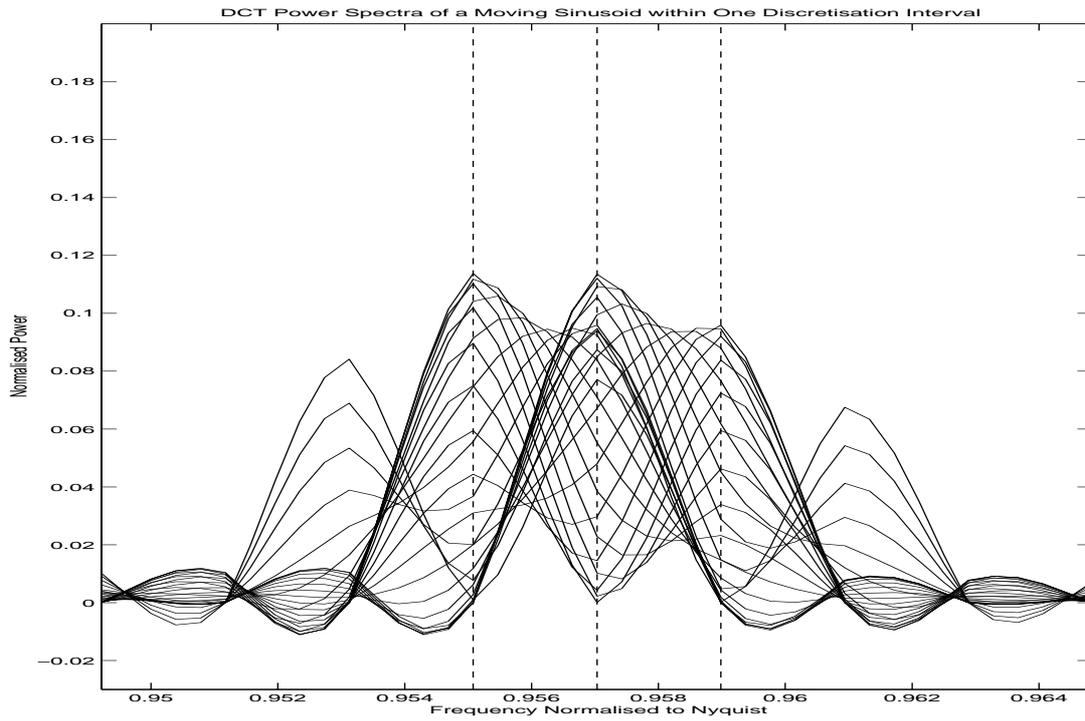


Fig. 7. DCT power spectra of a cosine signal whose frequency (slightly below the Nyquist frequency) changes within Δf . This figure clearly shows a rather poor frequency resolution due to the phase shift of the DCT basis function.

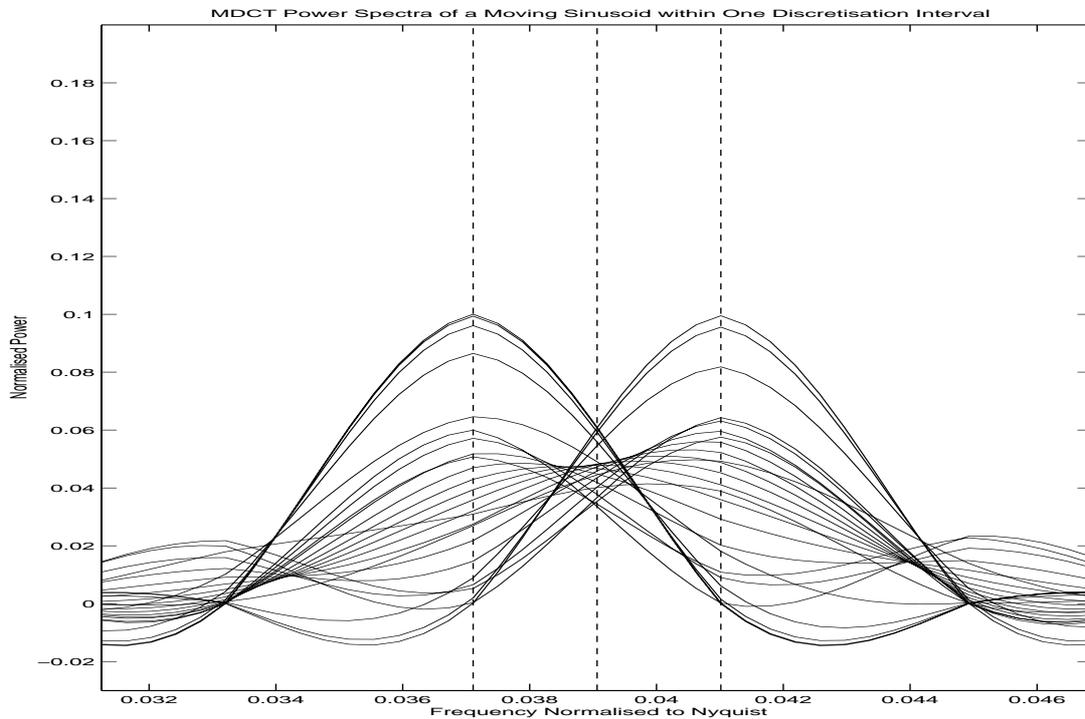


Fig. 9. MDCT power spectra of a cosine signal whose frequency (in the low frequency range) changes within Δf .

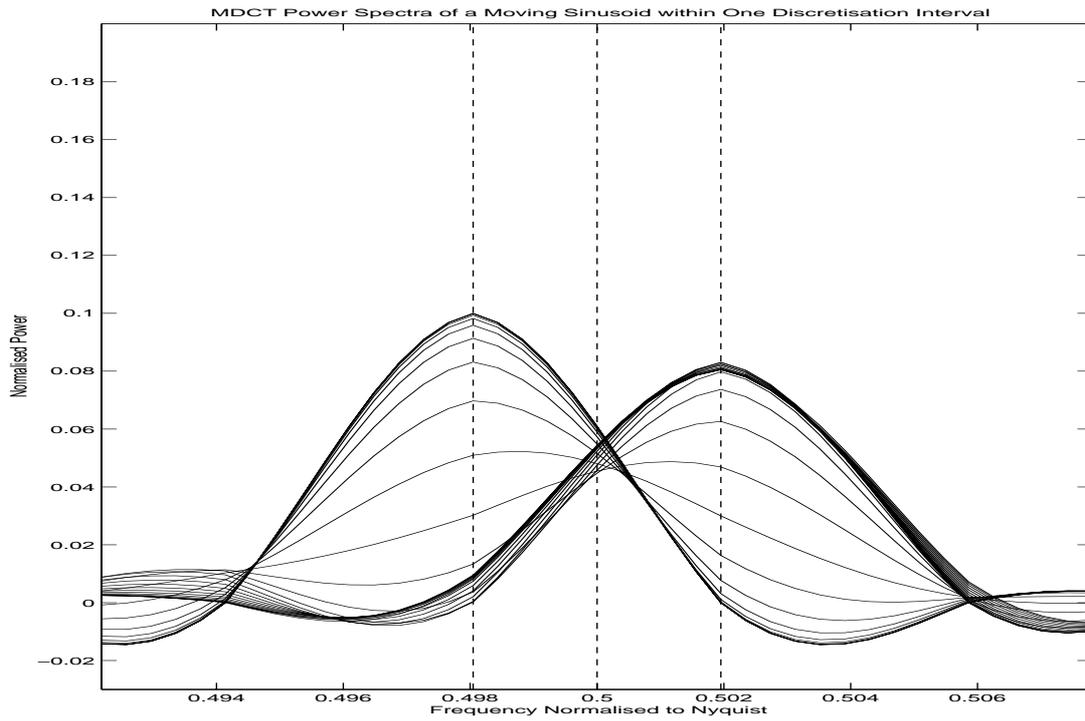


Fig. 10. MDCT power spectra of a cosine signal whose frequency (around half of the Nyquist frequency) changes within Δf .

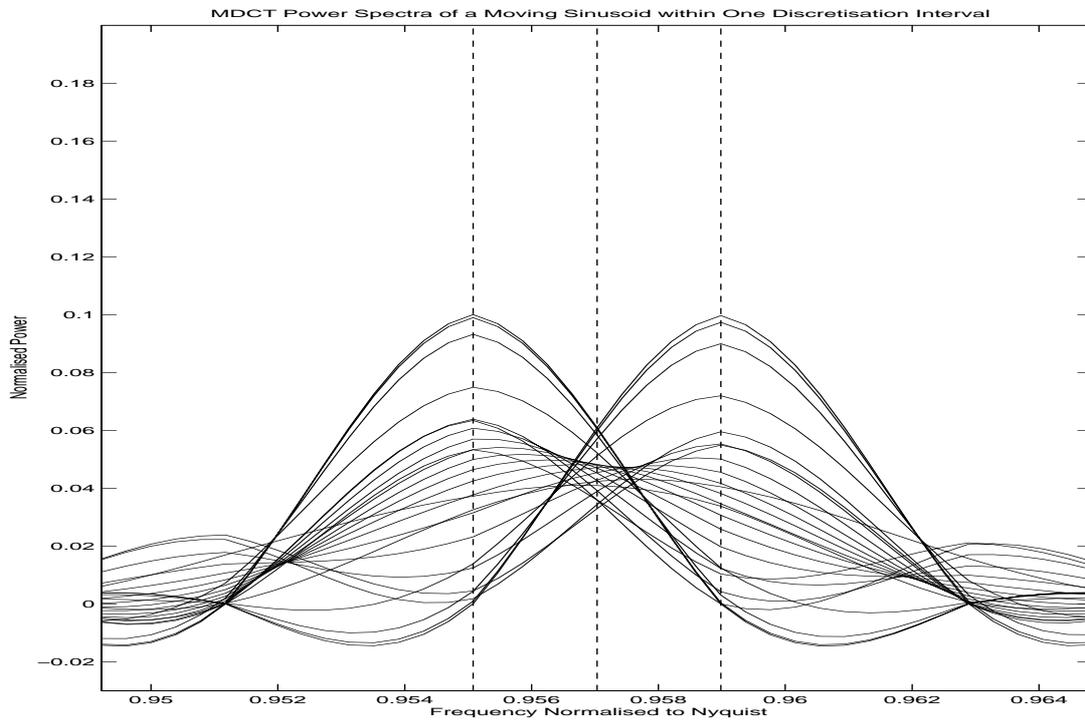


Fig. 11. MDCT power spectra of a cosine signal whose frequency (slightly below the Nyquist frequency) changes within Δf .

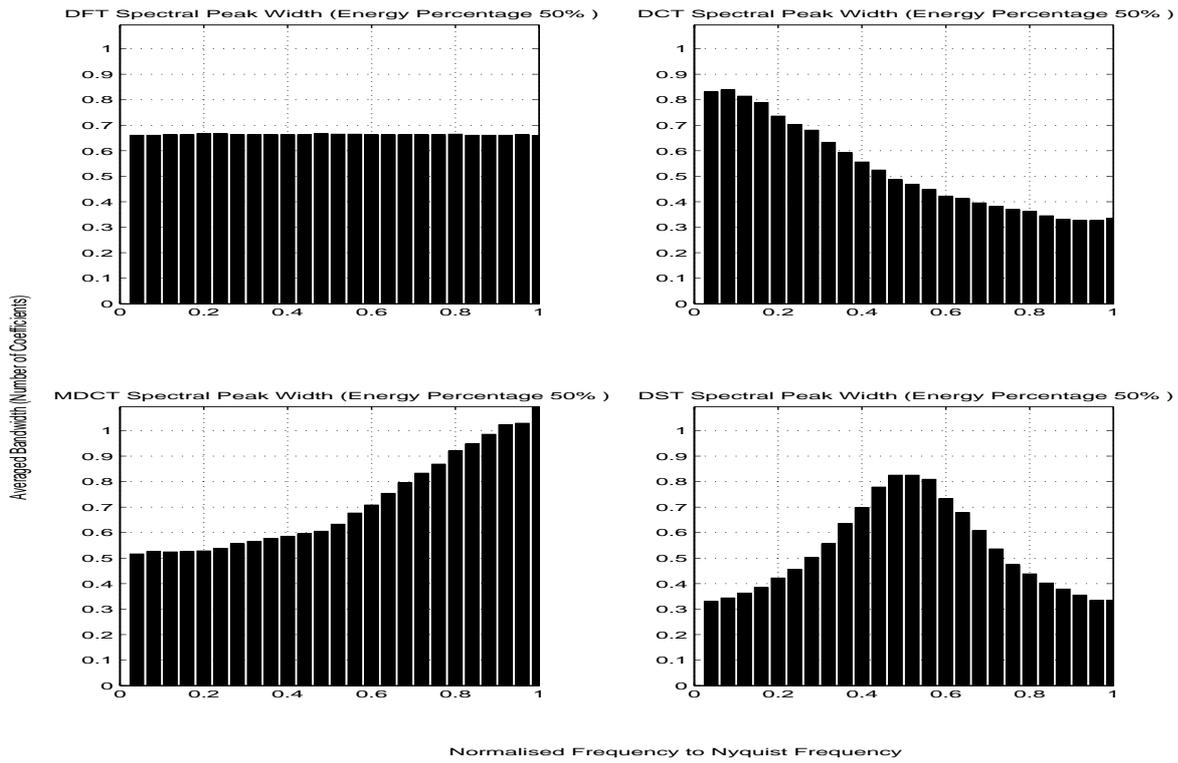


Fig. 12. Comparison of spectral resolutions (spectral peak width) of DFT, DCT, MDCT and DST using sine signals of 512 samples (rectangular window) as a function of signal frequency. The normalized power spectral threshold is 0.5, which corresponds to 50% energy within the spectral peak width.

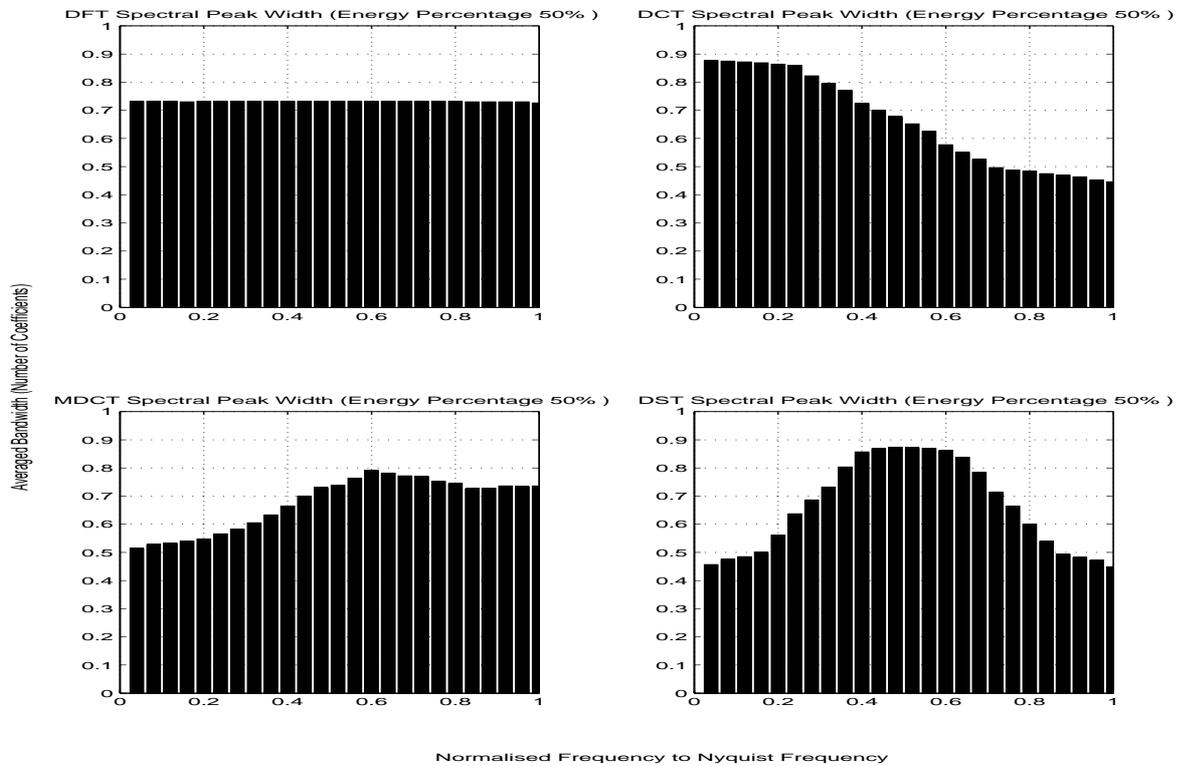


Fig. 13. Comparison of spectral resolutions of DFT, DCT, MDCT and DST using sine signals of 512 samples (sine window) as a function of signal frequency.

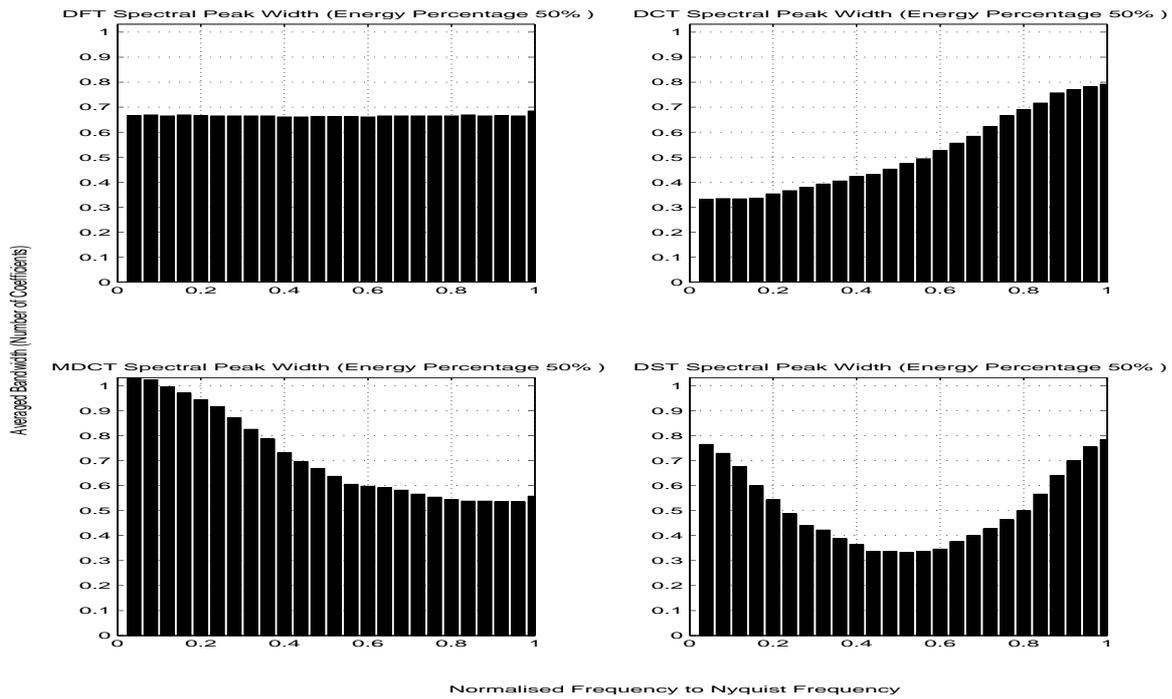


Fig. 14. Comparison of spectral resolutions (spectral peak width) of DFT, DCT, MDCT and DST using cosine signals of 512 samples (rectangular window) as a function of signal frequency.

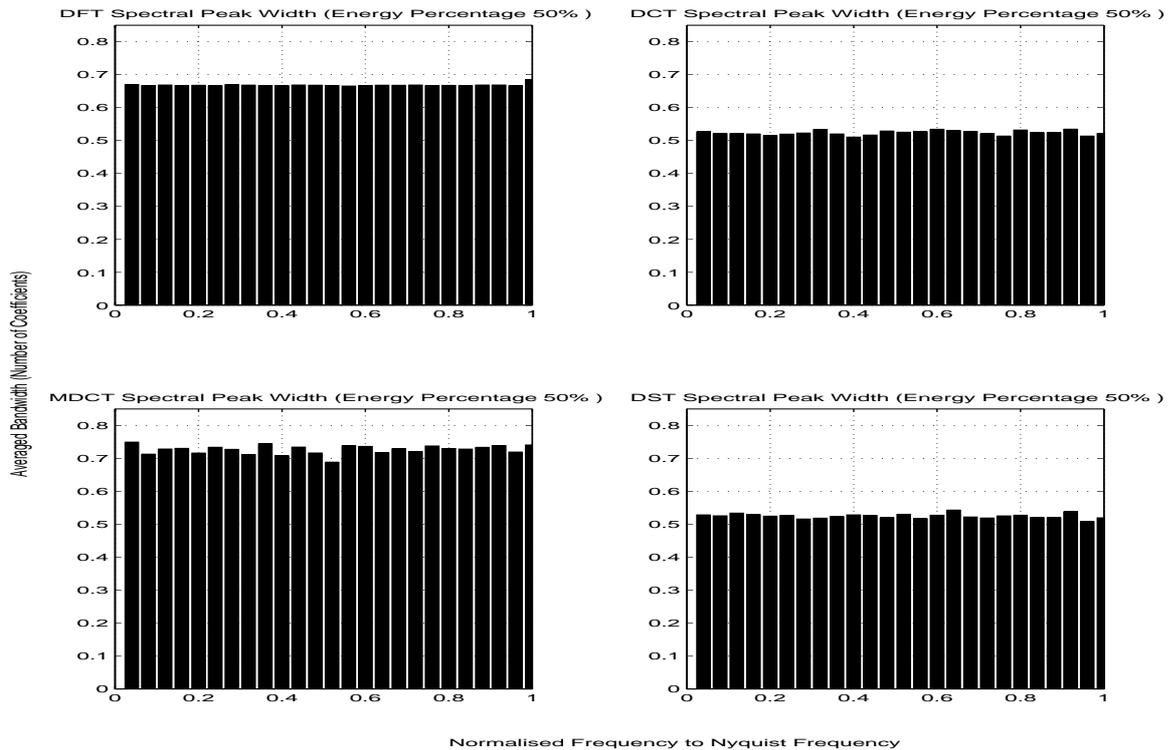


Fig. 15. Comparison of spectral resolutions (spectral peak width) of DFT, DCT, MDCT and DST using cosine signals with random phase shifts (rectangular window length = 512 samples) as a function of signal frequency.

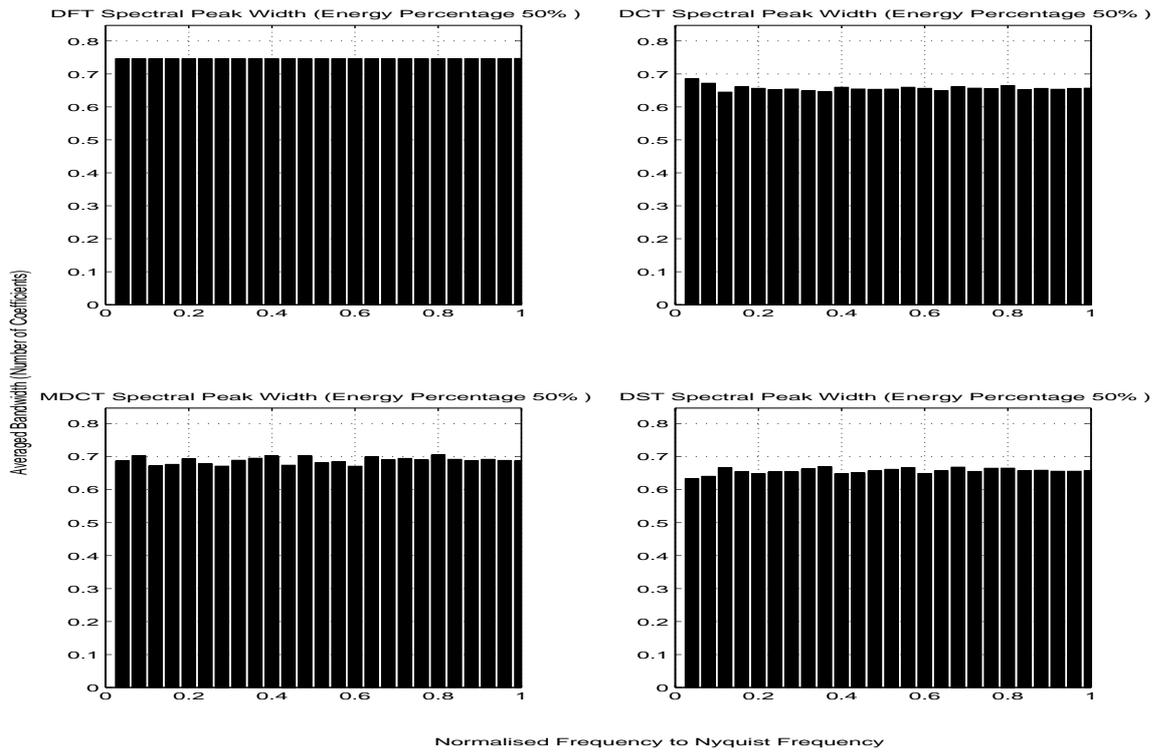


Fig. 16. Comparison of spectral resolutions (spectral peak width) of DFT, DCT, MDCT and DST using cosine signals with random phase shifts (sine window length = 512 samples) as a function of signal frequency.

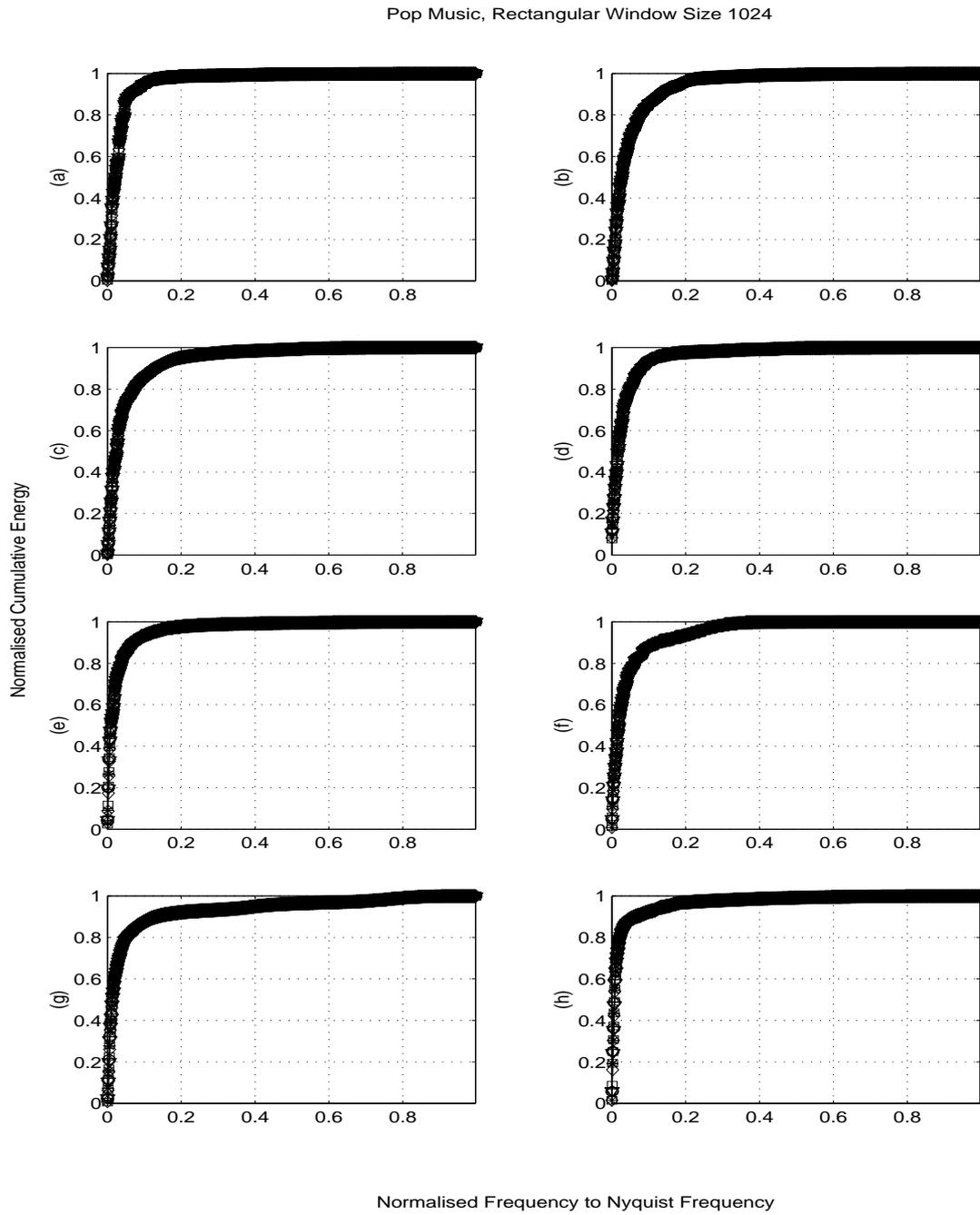


Fig. 17. Comparison of energy compaction property of DFT (star), $SDFT_{\frac{N+1}{2}, \frac{1}{2}}$ (circle), DCT (diamond), MDCT (triangle) and DST (square) using 8 pieces of pop music with a rectangular window size = 1024 samples. In this scale, it is impossible to distinguish the difference between different transforms.

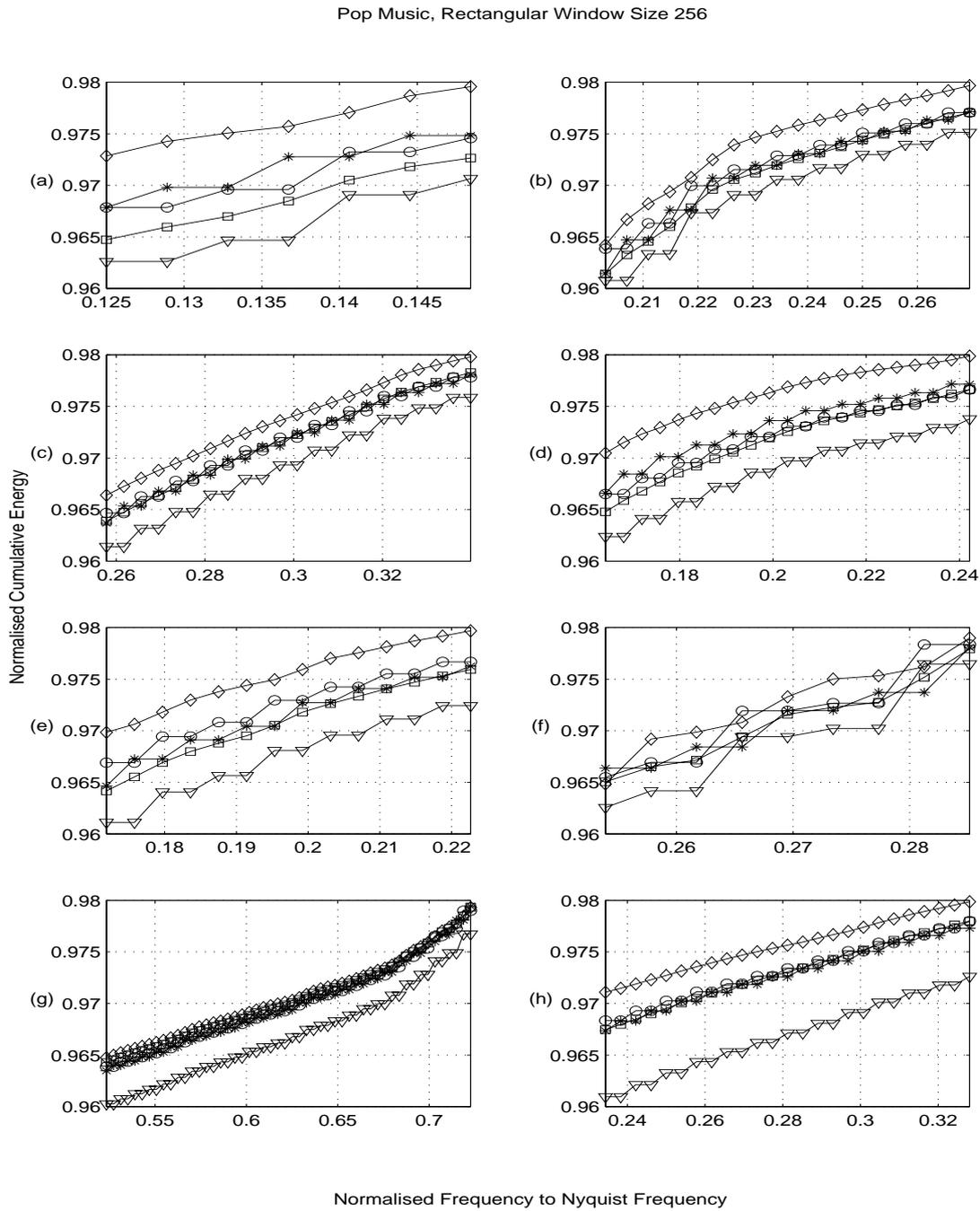


Fig. 18. Comparison of energy compaction property of DFT (star), $SDFT_{\frac{N+1}{2} \times \frac{1}{2}}$ (circle), DCT (diamond), MDCT (triangle) and DST (square) using 8 pieces of pop music with a rectangular window size = 256 samples. This is a zoomed version for better illustration.

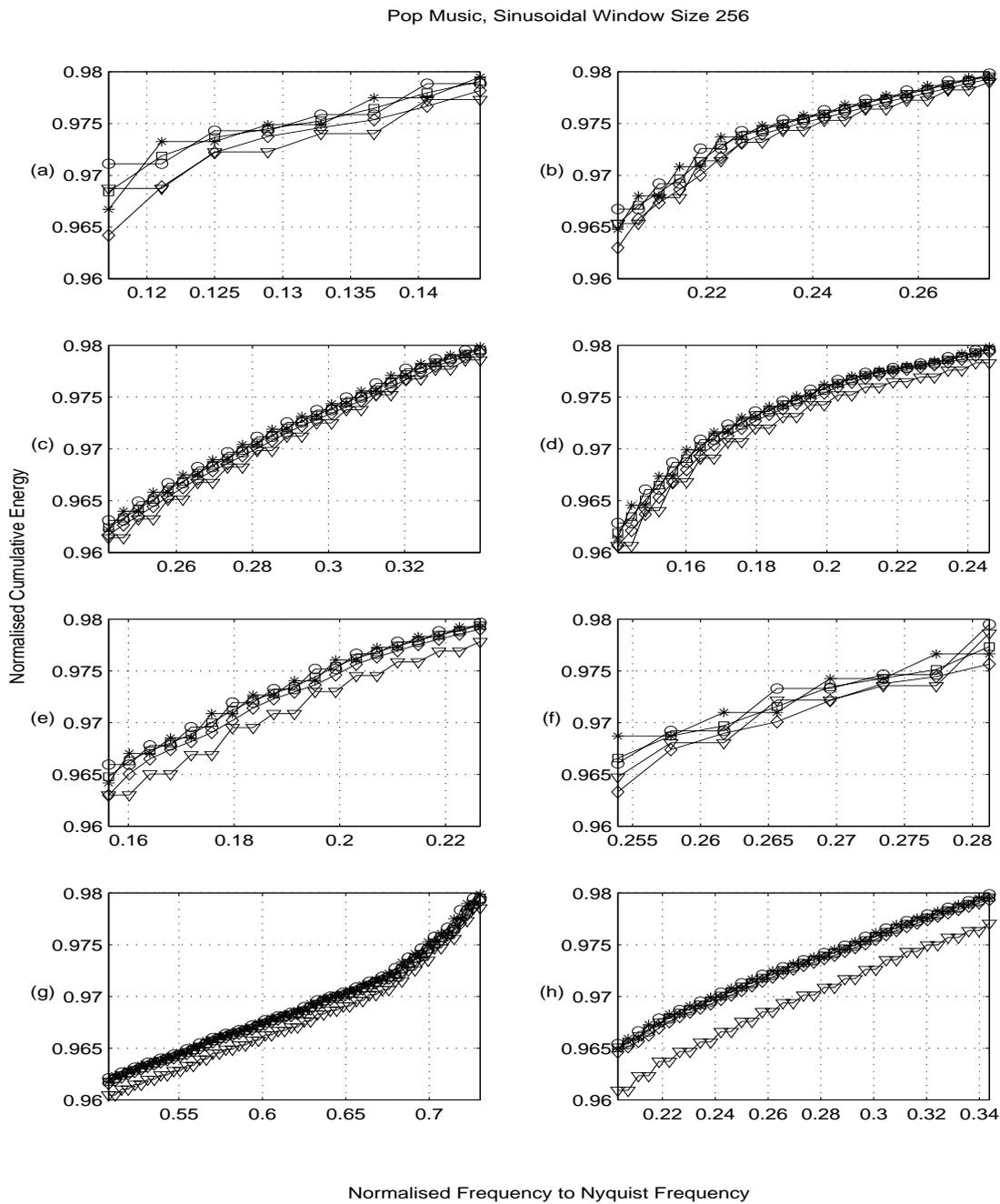


Fig. 19. Comparison of energy compaction property of DFT (star), $SDFT_{\frac{N+1}{2}, \frac{1}{2}}$ (circle), DCT (diamond), MDCT (triangle) and DST (square) using 8 pieces of pop music with a sine window size = 256 samples. The sine window clearly reduces the difference between different transforms.

Classical Music, Rectangular Window Size 1024

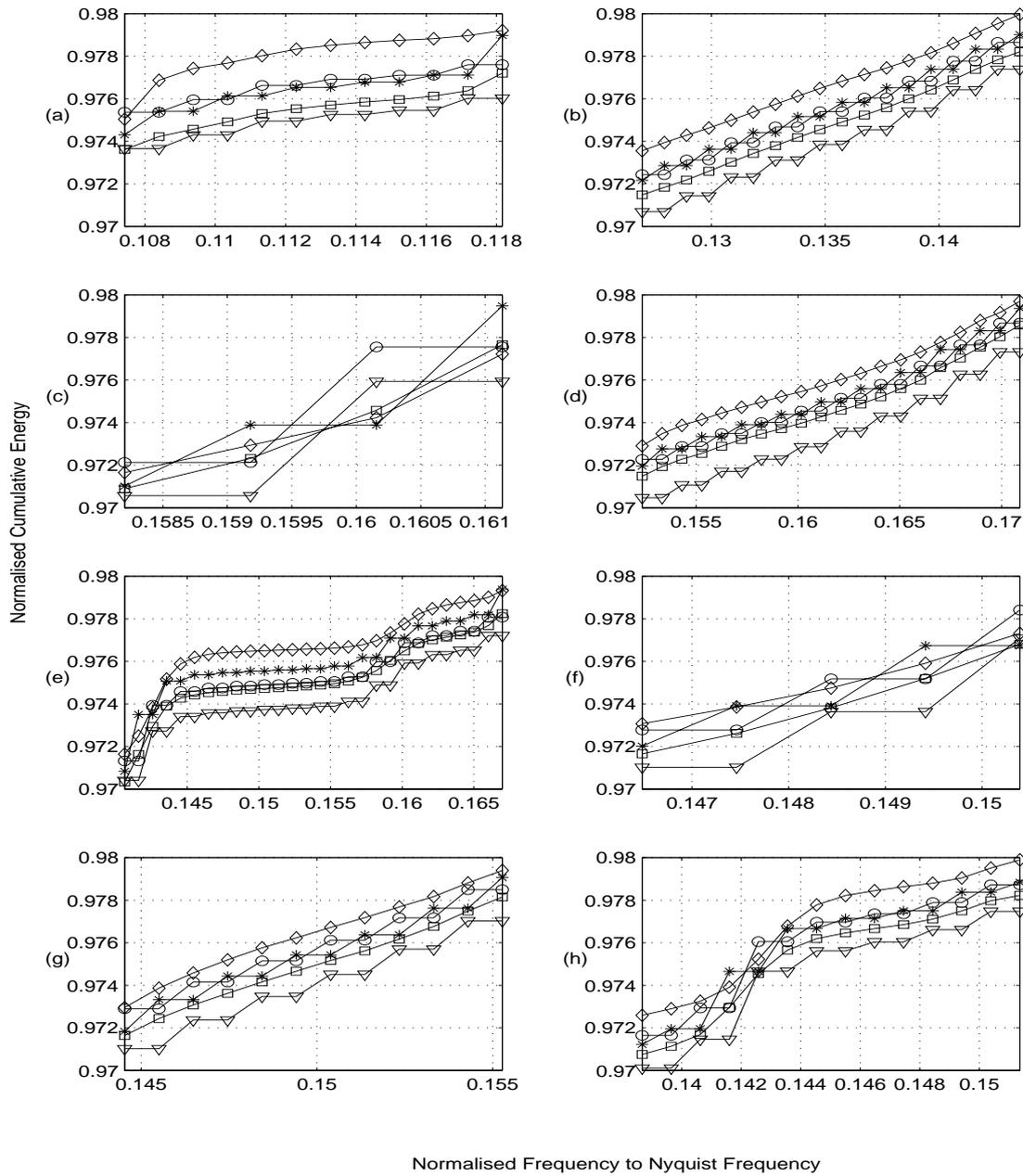


Fig. 20. Comparison of energy compaction property of DFT (star), $SDFT_{\frac{N+1}{2}, \frac{N-1}{2}}$ (circle), DCT (diamond), MDCT (triangle) and DST (square) using 8 pieces of classic music with a rectangular window size = 1024 samples

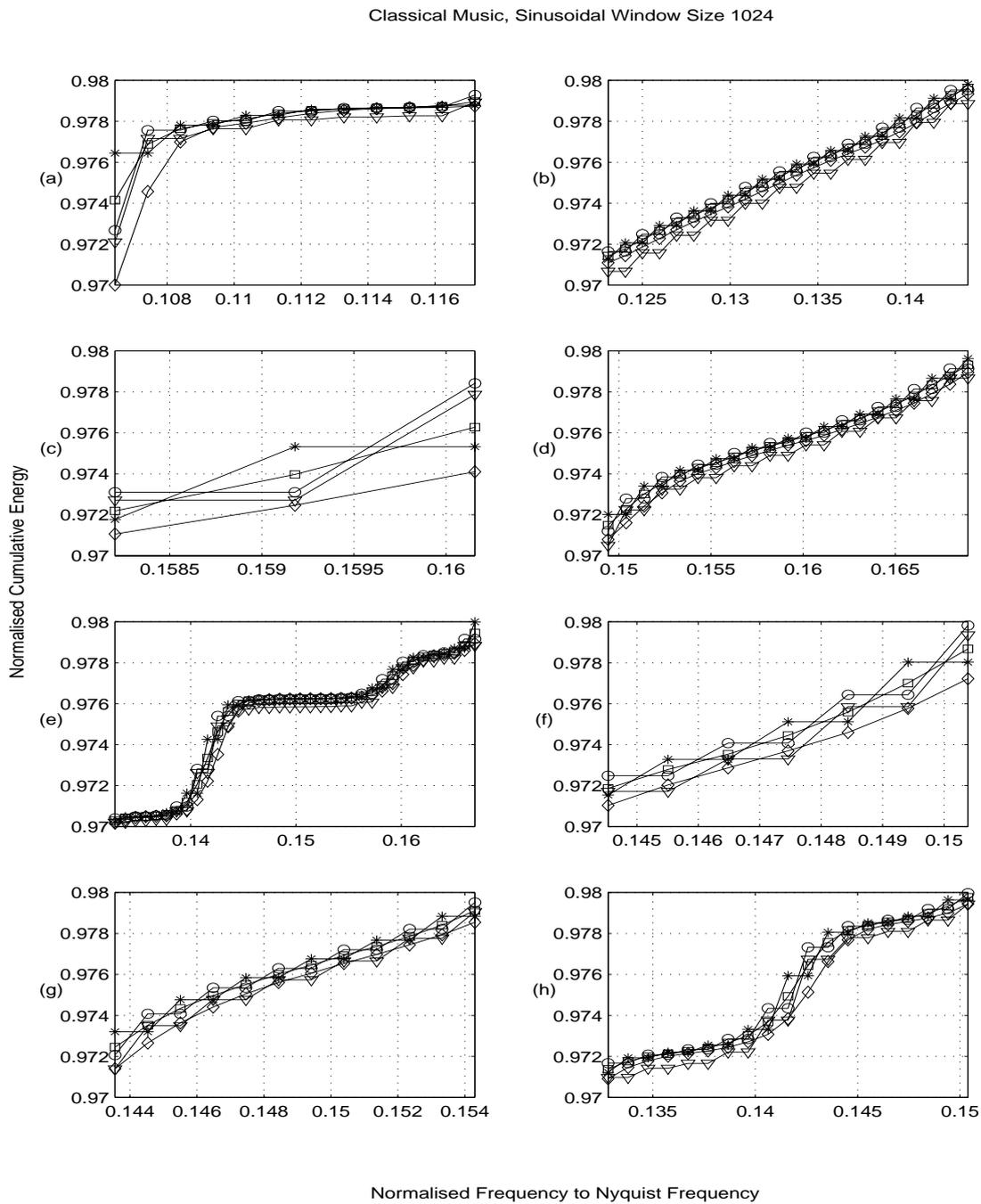


Fig. 21. Comparison of energy compaction property of DFT (star), $SDFT_{\frac{N+1}{2}, \frac{1}{2}}$ (circle), DCT (diamond), MDCT (triangle) and DST (square) using 8 pieces of classic music with a sine window size = 1024 samples

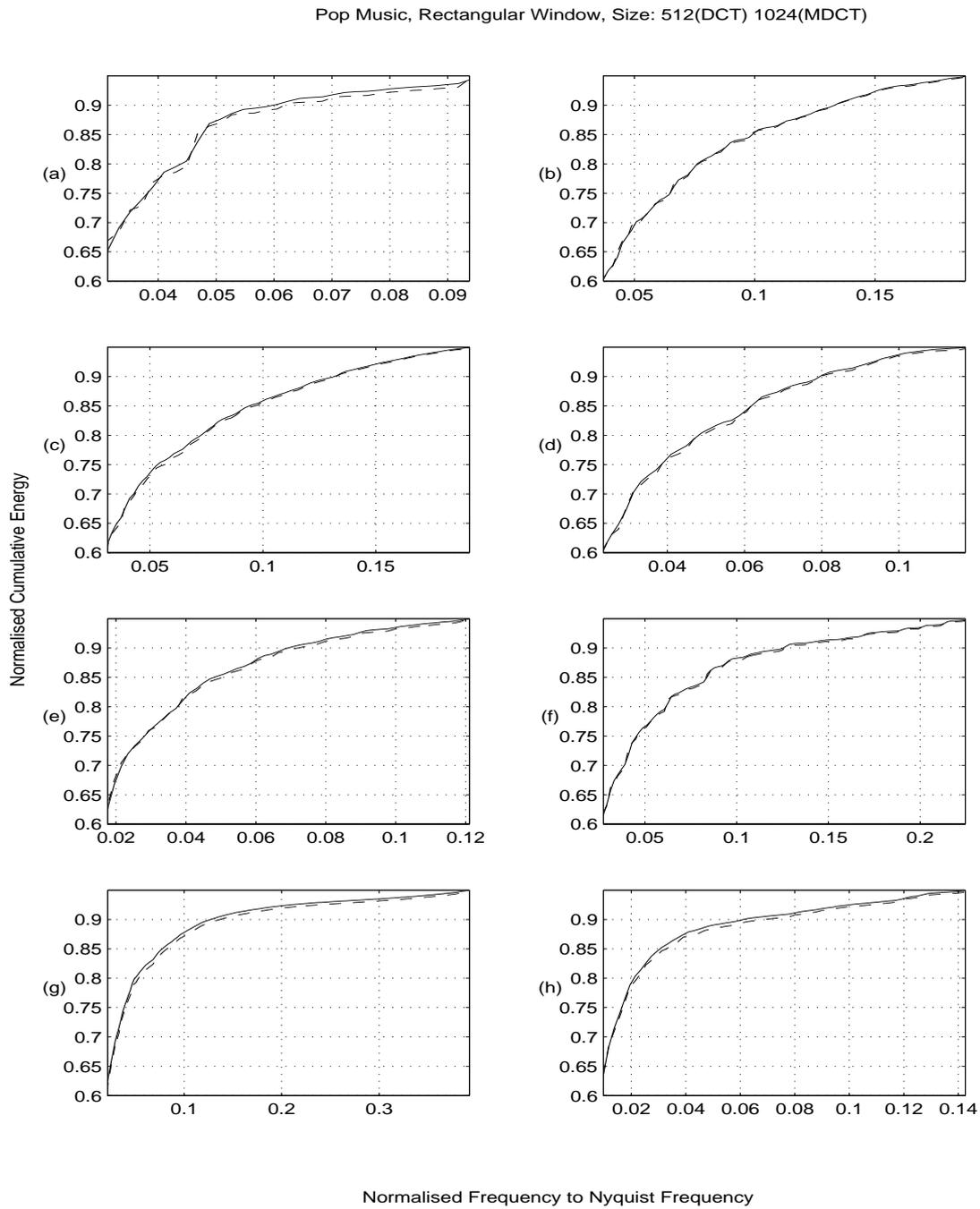


Fig. 22. Comparison of energy compaction property using 8 pieces of pop music with a rectangular window size of 512 samples for DCT (solid lines) and a rectangular window size of 1024 samples for MDCT (dashed lines).

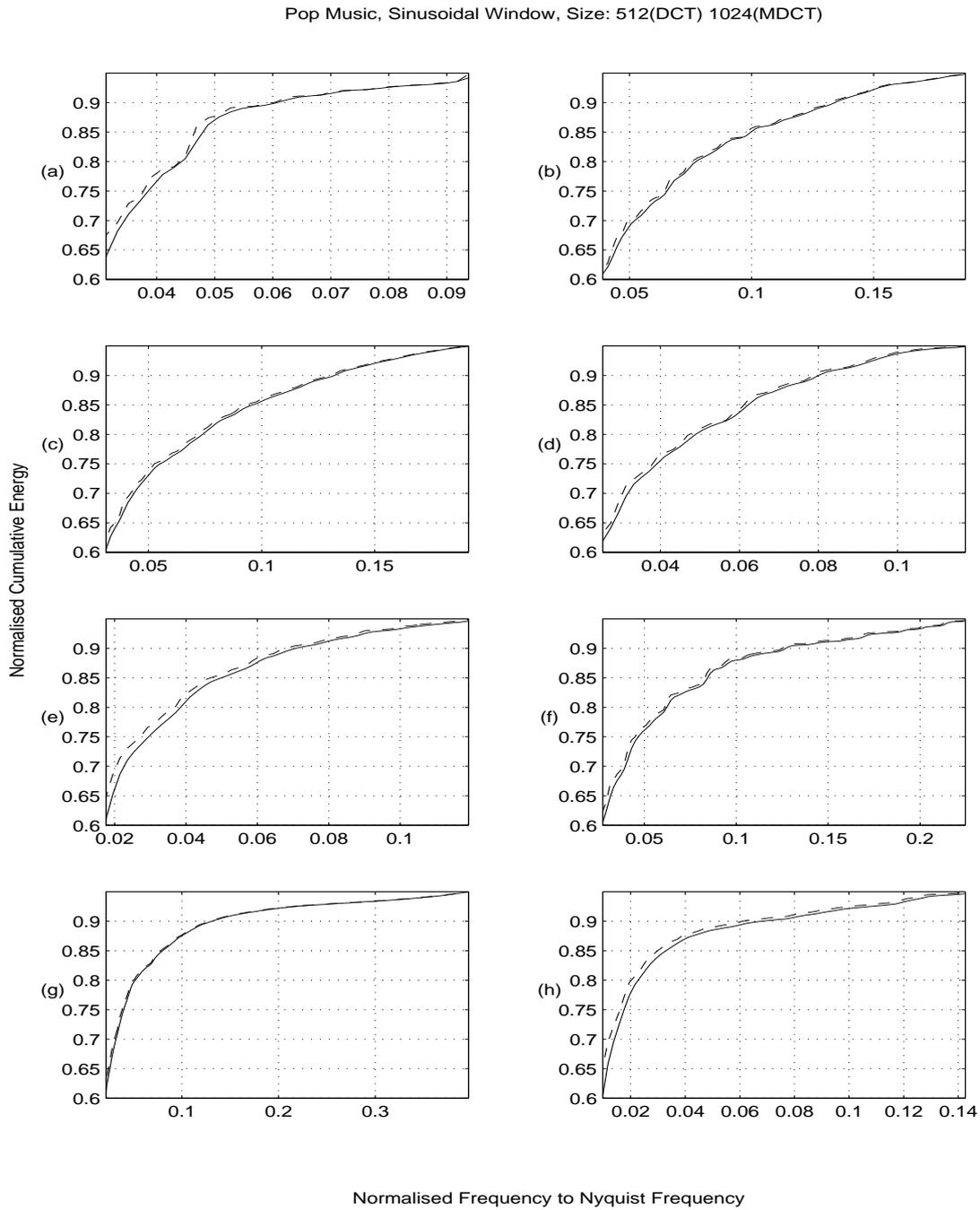


Fig. 23. Comparison of energy compaction property using 8 pieces of pop music with a sine window size of 512 samples for DCT (solid lines) and a sine window size of 1024 samples for MDCT (dashed lines).

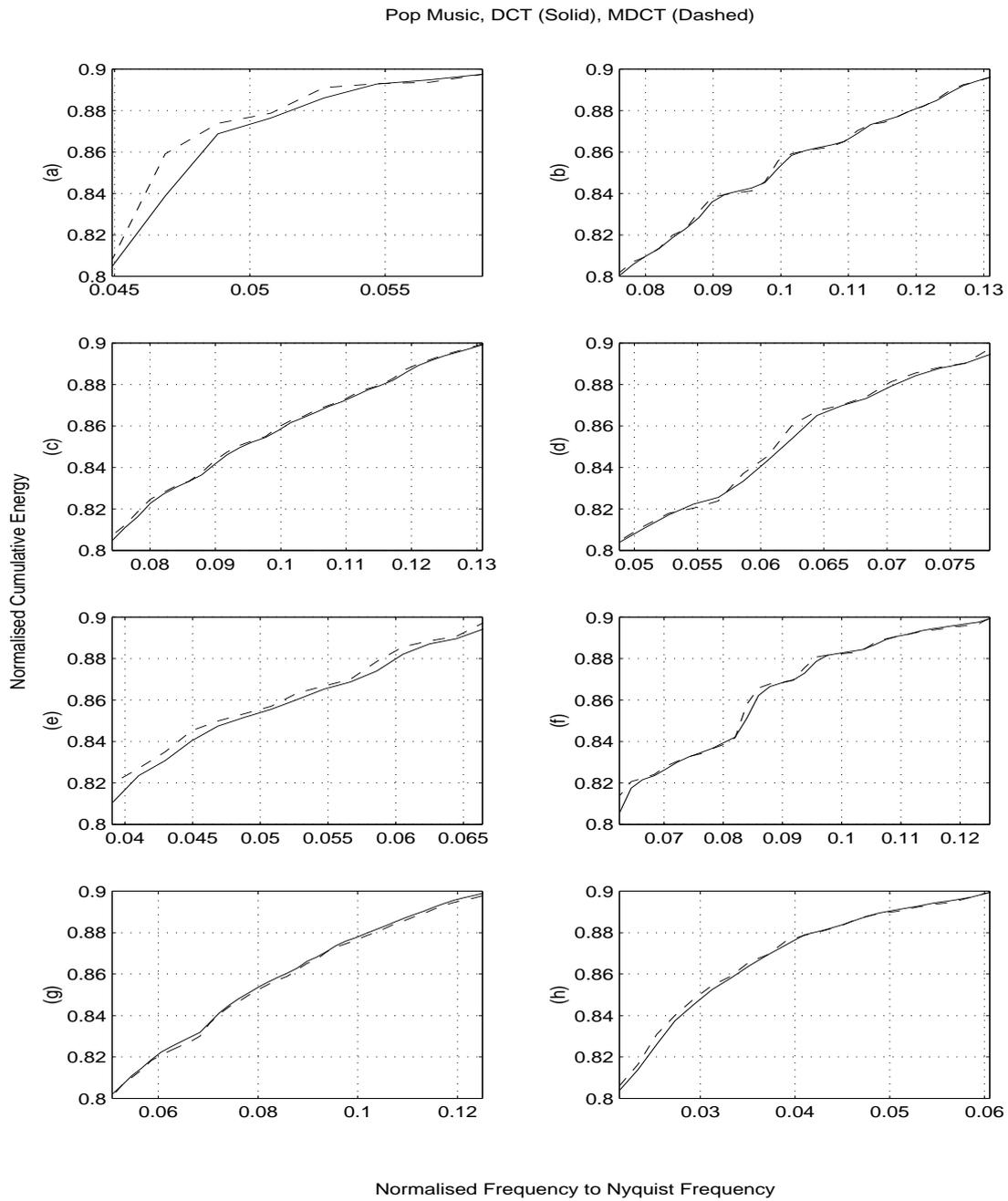


Fig. 24. Comparison of energy compaction property using 8 pieces of pop music with a rectangular window size of 512 samples for DCT (solid lines) and a sine window size of 1024 samples for MDCT (dashed lines).

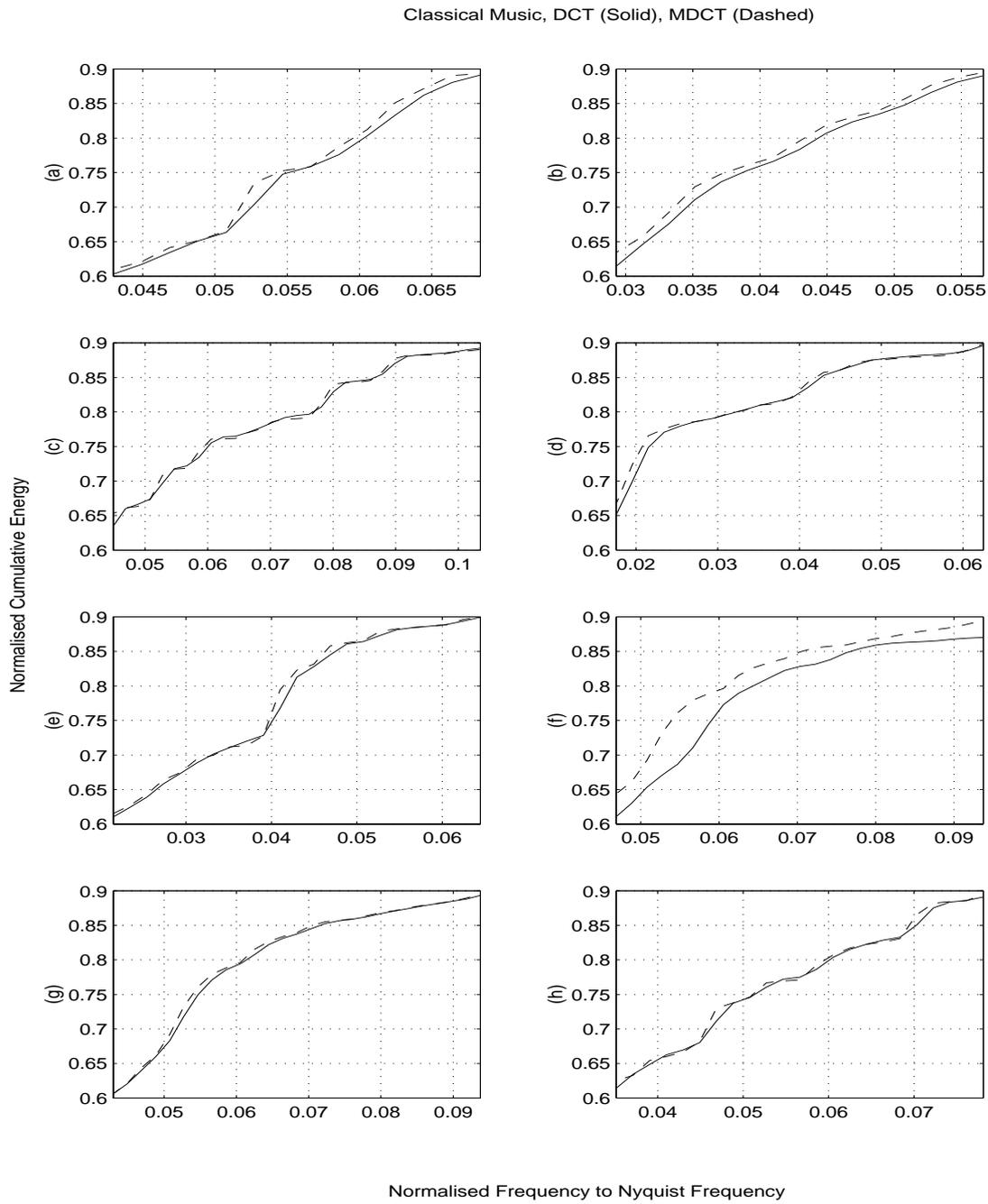


Fig. 25. Comparison of energy compaction property using 8 pieces of classic music with a rectangular window size of 512 samples for DCT (solid lines) and a sine window size of 1024 samples for MDCT (dashed lines).