FREQUENCY AVERAGING:
AN USEFUL MULTIWINDOW SPECTRAL ANALYSIS APPROACH

Climent Nadeu, Jaume Padrell and Ignasi Esquerra

Universitat Politècnica de Catalunya, Barcelona, Spain

ABSTRACT

The multiwindow approach is a meaningful framework for nonparametric spectral estimation. It also encompasses several conventional methods as WOSA and frequency-averaged periodogram. Recently, some authors claimed that the Slepian windows of the Thomson's method and other related optimal sets of windows show a better performance in terms of resolution, variance and leakage. In this paper, that claim is discussed by means of some simulation examples and by applying the various methods to speech recognition. In conclusion, frequency averaging of the periodogram is a computationally simple method that has a great flexibility for band specification and comparatively shows good performance. In fact, it is the spectral analysis technique most extensively employed for speech recognition.

1. INTRODUCTION

Since Thomson's work [1], several recent spectral analysis methods are based on the multiwindow (MW) approach. Given a signal $x(n)$ between $n=0$ and $n=N-1$, they estimate the power spectral density by averaging the windowed periodograms that result from K orthonormal windows or tapers $v_k(n)$, $0 \leq k \leq K-1$, $0 \leq n \leq N-1$, which are optimal in a given way.

The set of Slepian windows or discrete prolate spheroidal sequences used in the Thomson's method can be described as arising from the Karhunen-Loève eigenequation that, written in the frequency domain, is [1]

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} Q(\theta)K_N(\omega-\theta)V_k(\theta)d\theta = \lambda_k V_k(\omega) \qquad (1)$$

where $0 \leq k \leq K-1$, $V_k(\omega)$ is the Fourier transform of $v_k(n)$, λ_k is the corresponding eigenvalue, and $K_N(\omega)$ is the Dirichlet kernel, i.e. the Fourier transform of a rectangular window ranging from 0 to $N-1$.

The Slepian windows are obtained from the Karhunen-Loève eigenequation when $Q(w)$ of (1) is

$$Q(\omega) = \begin{array}{ll} 1, & -W < \omega < W \\ 0, & |\omega| > W \end{array} \qquad (2)$$

Choosing shapes for $Q(\omega)$ others than the rectangular one, different families of windows follow from the integral equations (1). Those orthonormal windows are also orthogonal with respect to $Q(\omega)$ as weight.

Every MW estimator computes an estimate $\hat{S}(\omega)$ of $S(\omega)$ by averaging in some way the power within a band surrounding the current frequency ω. Each window contributes to this average favouring some subbands in front of the others. If we wish to control that contribution, we have to assign different weights a_k to each windowed periodogram, i.e.

$$\hat{S}(\omega) = \sum_{k=0}^{K-1} a_k \left| \sum_{n=0}^{N-1} v_k(n)x(n)e^{-j\omega n} \right|^2 \qquad (3)$$

so that $\sum_{k=0}^{K-1} a_k = 1$.

In order to have a measure of the combined effect of the set of windows on the frequency domain, a *composite spectral window* (CSW) can be defined as

$$W(\omega) = \sum_{k=0}^{K-1} a_k |V_k(\omega)|^2 \qquad (4)$$

As shown in the Mullis-Scharf's tutorial of quadratic estimators [2], both the time averaging of periodograms (also called WOSA or Welch technique) and the frequency averaging of the periodogram (FAP) are also MW variants. In the WOSA case, the various windows are time-shifted versions of a base window. The FAP technique computes a weighted average of K periodogram values within a given band around the current frequency. Interpreting the FAP as a MW

technique, the K windows result from multiplying a base window $v(n)$ by complex exponentials that produce frequency shifts of its Fourier transform.

In this paper, some MW spectral estimators arising from the eigenfunction framework (1) are experimentally compared with the conventional MW estimators WOSA and FAP. After discussing in Section 2 two reported performance comparisons [3,4], we present in Section 3 some experimental examples that illustrate how the Thomson's method does not necessarily offer better statistic performance than the conventional MW methods, at least when both frequency and time resolution are controlled.

In the MW estimators based on (1), frequency resolution implicitly depends on the spectral function $Q(\omega)$. In the Thomson's approach, $Q(\omega)$ is rectangular (see (2)). As noted in [5], for peaked spectra, a sharper $Q(\omega)$ is able to obtain spectral estimates that show less bias at peaks, as we will see in Section 4.

Time resolution is relevant for estimating evolutionary spectra [6]. For instance, spectra of speech signals are estimated frame-to-frame, since it is assumed that every speech frame is approximately stationary. In Section 5, we compare the various methods for speech spectral estimation.

2. REMARKS ABOUT THE PERFORMANCE OF MW METHODS

Some already published works [3,4] that compare MW methods that use orthogonal windows (the Thomson's method [1] and the Clark-Mullis's method [4]) with the WOSA method arrive to the conclusion that WOSA can not achieve a performance so good as that of the optimal MW methods. However, those performance comparisons do not effectively control either the frequency or the time resolution of the various estimators.

In [3], the performance is measured in terms of frequency resolution, variance and leakage. According to that paper, the Thomson's MW method "always performs better that WOSA" since, selecting the design parameters so that two measures are the same for each estimator, the Thomson's method shows better results for the third one. In that work, the way of equalling the frequency resolution for both methods consists of using the same value of the bandwidth parameter W in (2) for the set of Slepian windows of the Thomson's method and for the only WOSA window, which is taken as a first Slepian sequence. However, the effective bandwidth of the CSW increases with the number of sequences K so that the set of Slepian windows has a broader CSW.

On the other hand, that work is not taking into account the fact that, unlike the Thomson's MW method, WOSA with non-rectangular window weights both sides of the data segment less than the middle part. Actually, it is assumed, like in Thomson's works, that only a segment of data is known and, consequently, attention is not payed to the higher resolution in the time domain that its WOSA

version can offer to estimate an evolutionary spectrum due to its smaller effective time length.

In the example presented in [4], the effective time length is taken into account by allowing an enlargement of the data segment for the WOSA method. In fact, as a triangular window is used for the WOSA estimator, a rectangular window of length N results from adding up the various shifted versions of the triangular window.

However, in [4], the authors do not intend to control the frequency resolution as they do not use the same effective bandwidth for all the methods. As they actually point out, whereas the CSW of the Thomson's method and that of their own method are roughly square-shaped, the magnitude of the transform of the triangular window has the main lobe of the squared sinc function. Hence, their WOSA estimator could show a larger frequency resolution. Unfortunately, it can not be observed in their simulations since they use as example the spectrum of an MA process which is rather smooth so that resolution is not of importance in it.

3. SOME ILLUSTRATIVE EXAMPLES

Using the same MA process as in [4], and the same number of data points, $N=250$, we have computed by Monte-Carlo simulations the means and the variances of the various techniques, intending to use the same frequency resolution for all of them. To measure the frequency resolution, we consider that it is inversely proportional to the Parzen's effective bandwidth [6] of the CSW. Note that the WOSA's CSW is the only spectral window itself.

Thus, the triangular window of the WOSA method has been shortened with respect to [4] in order to have the same effective bandwidth as the Thomson's estimator. This shorter window allows a larger number ($K=7$) of time-shifted windows running along the 250 data points. Nevertheless, the effective time length of the WOSA estimator is still smaller than that of the Thomson's one due to the smooth edges of the set of time-shifted windows at both ends of the data segment. Notice that up to 300 points were used in [4] for the WOSA method in order to have an effective time length similar to that of the Thomson's method. There, the windows were shifted half their length; in our experiment, the shifting is just slightly larger.

Unlike in [4], the FAP method was also compared in our simulation tests with the above two methods. Five periodogram points were averaged at every frequency (i.e. $K=5$) in order to have an effective bandwidth approximately equal to that of the Thomson's method. A basic rectangular window $v(n)$ was used over the entire data segment for the FAP method to have approximately the same effective time length than the Thomson's method. Correspondingly, the K spectral windows are shifted versions of $(K_N(\omega))^2$. Usually, the weights a_k in (3) are constant for the Thomson's estimator. They were taken constant for both methods in the simulations.

3954
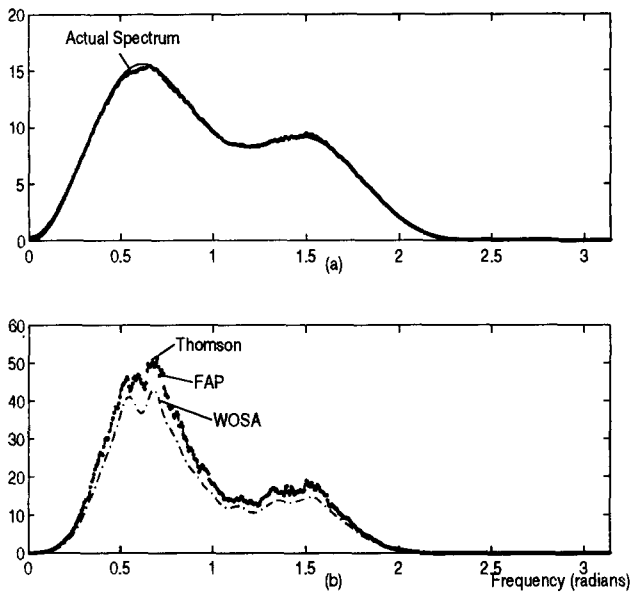
Fig.1 MA process: (a) Sample means. (b) Zoom of the peaks area. (c) Sample variances.

The estimated means and variances of the three methods for that MA process are depicted in Fig.1. As apparent, the mean curves of the three estimators are very similar and close to the exact spectrum, and the variance curves of the Thomson's and the FAP methods look like the same. Additionally, the variance for WOSA is not higher than that of the other two methods, but rather lower. An opposite observation about WOSA's variance was made in [4], in spite of the larger number of data (300 points) that were used there. This difference can be explained by the larger number of windows that can be used in our WOSA version due to the fact that its frequency resolution is made equivalent to those of the other methods.

We next performed some tests with the peaky spectrum of the fourth-order already AR process used in [7], in order to compare more properly resolution and leakage of the various estimators. Fig.2 shows the sample means and variances of the spectrum estimates for the three methods, using the same amount of data (N=250) and the same number and type of windows that in the MA process.

Again, the mean and variance curves of the Thomson's and the FAP methods are almost coincident in the peaks area, and are close in the low-power intervals. In this AR case, the variance of WOSA is not larger than that of the Thomson's method neither. The WOSA mean shows a smaller leakage in the low-power regions due to the relatively low sidelobes of the triangular window.

When a rectangular window is used, the WOSA leakage is even larger than that of the other two methods. However, its variance at the peaks is approximately the same as that of the other methods, in spite of using only 5 non-overlapping (50 points long) windows in this case.
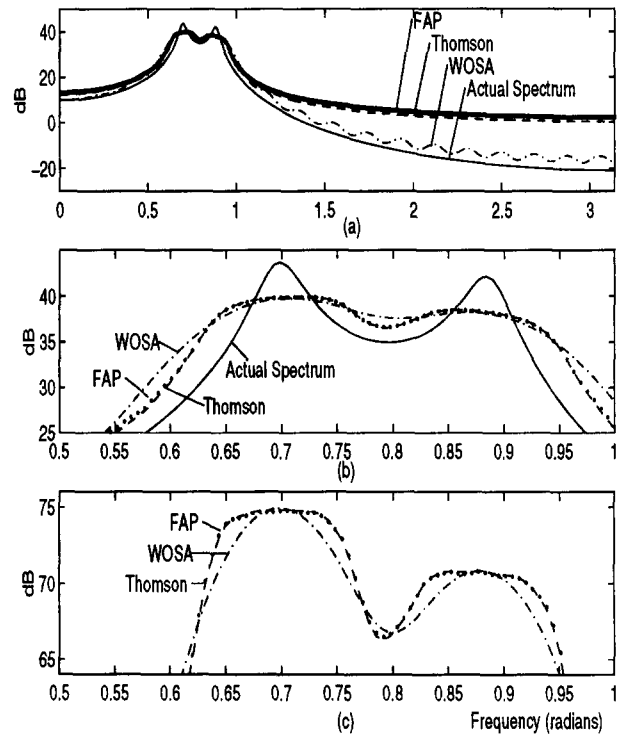
Fig.2 AR process: (a) Sample means. (b) Zoom of the peaks area. (c) Sample variances.

In conclusion, from our experiments, where both frequency and time resolution are approximately controlled, the Thomson's method does not seem to give better statistic performance than the conventional MW estimators. Actually, its results look like those of the FAP method.

4. NON-RECTANGULAR COMPOSITE SPECTRAL WINDOW

The WOSA method is more able to indicate the peaks position of the AR spectrum due to the fact that its main spectral lobe is sharper than the main lobe of the CSW of the Thomson's method or the FAP method, which are roughly square-shaped.

In order to obtain main lobes that are similarly shaped, we modified both FAP and Thomson's methods. The CSW of the former was easily controlled by using a weighted average of the periodogram points and adequately choosing the set of weights. To control the main lobe of the Thomson's CSW, $Q(\omega)$.in (1) was given a triangular form between $-W$ and W, and a_k were made equal to the corresponding eigenvalues λ_k normalized with respect to their summation. The same technique with a different shape for $Q(\omega)$ was already presented in [5]. In this way, a modified Thomson's method results whose set of windows shows an approximately triangular

$W(\omega)$. That CSW would be even closer to the triangle $Q(\omega)$ if N were larger.

These modified techniques were tested again with the same AR data. K was increased to 9 for them since the width of the main lobe of the CSW is approximately twice that of the non-modified methods. Interestingly enough, the windows of the modified Thomson's method are much more regularly distributed in frequency than the Slepian windows and their spectral shapes closely ressemble those of the FAP ones, except in they show two symmetric main lobes around the center frequency of the averaging band.

Fig.3 depictes the sample mean and variances of the three estimators. In accordance with the observation of the last paragraph about the shapes of the spectral windows, the means of the modified FAP and Thomson's methods are even closer than those of the non-modified methods. Additionally, the three mean curves almost coincide in the peak region. The variances are again similar.

Summarizing, from the experiments we observe that, when the CSW of the various techniques have a similar shape, the statistic performance of the estimators in terms of bias and variance is almost identical. Additionally, if we know a priori that the spectrum has prominent peaks, it will be useful to use a non-rectangular CSW.
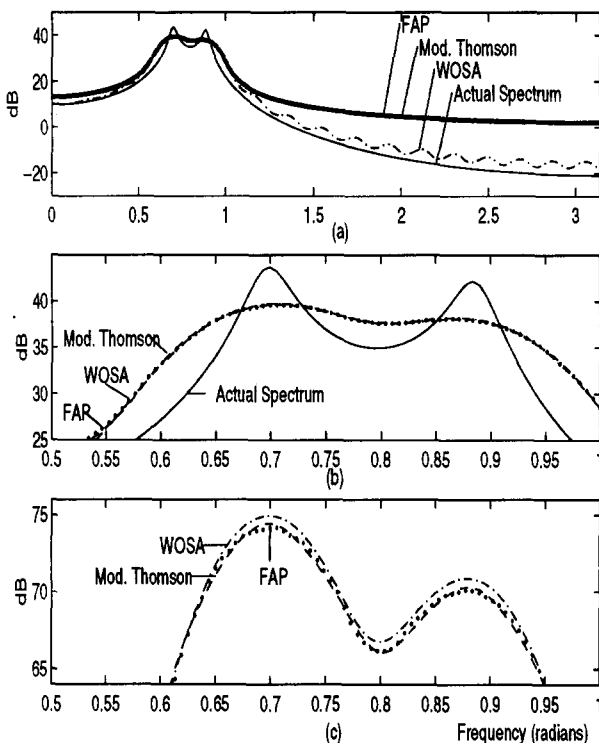
Fig.3 AR process and triangular-shaped CSW: (a) Sample means. (b) Zoom of the peaks area. (c) Sample variances.

5. APPLICATION TO SPEECH RECOGNITION

Assuming as usual that, in a short-term basis (20-30ms), the speech signal can be modelled by a stationary process, a frame-to-frame spectral analysis yields a temporal sequence of spectral estimates that represent the acoustic-perceptual content. As spectra of voiced speech contain peaks (the formants), a CSW with a peaky shape is a sensible choice for MW estimators. In fact, the spectral parameters that are most widely used for speech recognition (the mel-frequency cepstral coefficients) are obtained through spectral estimation with the FAP technique –on a non-uniform (mel) frequency scale– and applying a triangular set of weights a_k on the square magnitude of the DFT samples.

With the very same three MW techniques of the last section, we performed a speech recognition experiment with the TI connected digits data base and using continous density hidden Markov models. The sampling rate was 8 KHz, the frame length was 250 (like in Section 4), and the frame shift was 80. The estimated powers of 12 bands uniformly distributed between 0 and 4 KHz were employed to represent a frame of the speech signal. The relative differences in string recognition rate between the three methods were less than 1%. Hence, the result of this application is in accordance with the remarks arising from the above simulations.

6. CONCLUSION

The weighted FAP is a sensible choice for practical purposes since, apart from obtaining a performance similar to that of optimal methods like Thomson's one, it: 1) allows an easy control of the decisive composite spectral window, 2) shows a great flexibility in defining spectral bandwidths, and 3) has a relatively low computational load (only one actual window). In fact, with a triangular-like composite spectral window, a mel frequency scale, and Hamming-windowing the data to reduce leakage, FAP is the most employed method of spectral analysis for speech recognition.

REFERENCES

[1]   D. Thomson, *Proc. IEEE*, Vol. 70, No. 9, pp. 1055-96, Sept.1982.
[2]   C.T. Mullis, L.L.Scharf,, in *Adv. in Spectrum Analysis and Array Processing*, S. Haykin, Ed., Prentice-Hall, 1990.
[3]   T.P. Bronez, *IEEE Trans. SP*, Vol. 40, No. 12, pp. 2941-6, Dec. 1992.
[4]   M.P. Clark, C.T. Mullis, *IEEE Trans. SP*, Vol. 41, No. 1, pp. 222-31, Jan. 1993.
[5]   M. Hansson, T. Gänsler, Göran Salomonsson, *Proc. ICASSP'95*, Detroit, pp. 1617-20, May 1995.
[6]   M.B. Priestley, *Spectral Analysis & Time Series*, Acad. Press, 1981.
[7]   T.J. Ulrych, T.N. Bishop, in *Modern Spectrum Analysis*, D.G. Childers, Ed., IEEE Press, p. 54, 1978.