Technical University of Denmark

Evaluation Study of Fast Spectral Estimators Using In-vivo Data

Hansen, Kristoffer Lindskov; Gran, Fredrik; Pedersen, Mads Møller; Voxen, Iben Holfort; Jensen, Jørgen Arendt; Nielsen, Michael Bachmann

DTU Library
Technical Information Center of Denmark

Evaluation Study of Fast Spectral Estimators Using *In-vivo* Data

Hansen K.L.[1,2], Gran F.[2,3], Pedersen M.M.[1,2], Holfort I.K.[2], Jensen J.A.[2], Nielsen M.B.[1]

1) Department of Radiology, Section of Ultrasound, Rigshospitalet, Blegdamsvej 9, 2100 Kbh. Ø, Denmark
2) Center for Fast Ultrasound Imaging, DTU Elektro, Bldg. 349, Technical University of Denmark, 2800 Lyngby, Denmark
3) GN Resound, Lautrupbjerg 9, 2750 Ballerup, Denmark

Abstract – **Spectrograms in medical ultrasound are usually estimated with Welch's method (WM). To achieve sufficient spectral resolution and contrast, WM uses an observation window (OW) of up to 256 emissions per estimate. Two adaptive filterbank methods have been suggested to reduce the OW: Blood spectral Power Capon (BPC) and the Blood Amplitude and Phase EStimation method (BAPES). Ten volunteers were scanned over the carotid artery. From each dataset, 28 spectrograms were produced by combining four approaches (WM with a Hanning window (W.HAN), WM with a boxcar window (W.BOX), BPC and BAPES) and seven OWs (128, 64, 32, 16, 8, 4, 2). The full-width-at-half-maximum (FWHM) and the ratio between main and side lobe levels were calculated for each spectrogram. Furthermore, all 280 randomized spectrograms were presented to nine radiologists for visual evaluation: useful/not useful. BAPES and BPC compared to WM had better resolution (lower FWHM) for all OW<128 while only BAPES compared to WM had improved contrast (higher ratio). According to the scores given by the radiologists, BAPES, BPC and W.HAN performed equally well ($p>0.05$) at OW 128 and 64, while W.BOX scored less ($p<0.05$). At OW 32, BAPES and BPC performed better than WM ($p<0.0001$) and BAPES was significantly superior to BPC at OW 16 ($p=0.0002$) and 8 ($p<0.0001$). BPC at OW 32 ($p=0.29$) and BAPES at OW 16 ($p=0.55$) scored as at OW 128. WM at OW 16 and 8 failed as the four methods at OW 4 and 2. The intra-observer variability tested for three radiologist showed on average good agreement (90%, $\kappa=0.79$) and inter-observer variability showed moderate agreement (78%, $\kappa=0.56$). The overall result indicates that BPC and BAPES have better resolution and BAPES better contrast than WM, and that OW can be reduced to 32 using BPC and 16 using BAPES without reducing the usefulness of the spectrogram. This could potentially increase the temporal resolution of the spectrogram or the frame rate of the interleaved B-mode images.**

I. INTRODUCTION

Spectral Doppler in medical ultrasound is a quantitative technique providing estimates of blood velocity distribution within a range gate on a single image line. In most commercial scanners Welch's method (WM) [1] is used for the estimation. The axial velocity of the blood is found by imaging the same image line repeatedly, where WM uses up to 256 emissions per estimate [2]. The estimates are angle-corrected by the operator and presented as blood velocities plotted against time, denoted a spectrogram.

Conventional spectral estimation has limitations due to the long observation window (OW). The frame rate of the

accompanying B-mode images will be low [2] and the temporal resolution of the spectrogram will be impaired [3].

Alternative methods for fast velocity estimation have been proposed by several groups. In this paper the performance of two adaptive filterbank methods are investigated: Blood spectral Power Capon method (BPC) proposed by Stoica et al. [4] and Blood spectral Amplitude and Phase EStimation (BAPES) proposed by Gran et al. [5].

The purpose of this paper is to evaluate the two adaptive filtering methods compared to the conventional WM using *in-vivo* data. Ten volunteers were scanned and from each data set, spectrograms were found by using BPC, BAPES and WM on a series of different OWs. The calculated spectrograms were evaluated quantitatively and qualitatively.

II. MATERIALS AND METHODS

A. Volunteers

This prospective study was performed after approval by The Danish National Committee on Biomedical Research Ethics. Ten healthy volunteers (nine males and one female, mean age: 29.1 years, range: 24-36 years) entered the study after informed consent.

B. Spectral estimators

WM is applied on data using two different weighting schemes to control contrast and resolution in the resulting spectrogram: WM with a Hanning window (W.HAN) and WM with a boxcar window (W.BOX). W.HAN is the preferred conventional spectral estimator for which the resulting spectrogram has improved contrast at the expense of spectral resolution. Spectrograms generated with the W.BOX approach have improved spectral resolution but decreased contrast.

BPC and BAPES are alternative methods using data-dependent adaptive filtering techniques based on a matched filterbank framework. For a given data set, an optimal filter is calculated for each velocity component to minimize noise and interference from other velocities, whereby a unique filterbank is generated. In BPC the filters are designed to minimize the total power of the filtered data, while in BAPES the filters are designed to minimize the power of the filtered noise. In general these methods display better velocity resolution than

traditional methods like W.HAN and W.BOX. Details of the implementation can be found in [5].

C. Setup

The volunteers were scanned on the right common carotid artery using the experimental scanner RASMUS [6] and a B-K Medical 8804, 7 MHz linear array transducer. The setup is provided in Table 1. One data set of 2.5 seconds was recorded for each volunteer, where the sequential data acquisition alternated between a sequence for flow estimation of 3800 consecutive emissions and a sequence for B-mode imaging of 65 consecutive emissions. The recorded data was post processed using the four different methods: W.HAN, W.BOX, BPC and BAPES. For each method, spectrograms were estimated using different OWs: 128, 64, 32, 16, 8, 4 and 2 emissions per estimate. Hence, 28 spectrograms were calculated from every data set giving in total 280 spectrograms for all volunteers. Post processing was done offline with MATLAB (Mathworks, Natick, MA, USA) on a 100 CPU Linux cluster.

TABLE I
Scanner and transducer setup

Parameter	Value	Unit
Sampling frequency	40	MHz
Center frequency	7	MHz
Pulse repetition frequency	9.3	kHz
Number of transmitting elements	64	elements
Apodization in transmit	Uniform	N/A
Number of receiving elements	64	elements
Apodization in receive	Hanning	N/A
Wall Motion Filter	Mean-subtraction	N/A

D. Quantitative evaluation

To investigate the methods quantitatively, a spectrum at end-diastole was found for each spectrogram. For each data set, the time instant was chosen by visually examining the spectrogram obtained with W.HAN at OW 128. For the selected time, spectra from all 28 spectrograms of the same data set were found. From each spectrum, two parameters were calculated: the FWHM and the ratio between main and side lobe levels. The FWHM was found as the width of the velocity distribution of the main lobe at half of the maximum amplitude. Thus, the FWHM, given in m/s, is a measure of spectral resolution and should be as low as possible. The main to side lobe ratio was found as the relative difference between the side-lobe level and the peak amplitude. The side-lobe level was found as the median value of the distributed amplitudes outside the main-lobe, outlined by the FWHM. The ratio, given in decibel (dB), is a measure of contrast in the spectrogram and should be as high as possible.

E. Qualitative evaluation

To investigate the methods qualitatively, nine experienced radiologists evaluated in a blinded trial the 280 randomized spectrograms by scoring each spectrogram: useful or not useful. Additionally, intra- and inter-observer variations were found. The intra-observer variability was assessed by comparing the scores given twice by three radiologists with more than 14 days between each session.

F. Statistics

A descriptive statistical analysis was computed on FWHM and ratio data. Scores given by the radiologists were pooled by method and window. Useful and not useful were coded with the dummy variables 1 and 0, and Bonferroni adjusted tests for multiple comparisons with $p<0.05$ considered significant were performed [7]. The intra- and inter-observer variability were found using Cohen's and Fleiss' kappa, respectively [8;9]. Statistical analyses were performed using SAS (SAS Institute, Cary, NC, USA) and MATLAB (Mathworks, Natick, MA, USA).

III. RESULTS

From the same data set with three systoles recorded, examples of spectrograms obtained with the four methods at OW 128 and 16, are presented (Fig. 1). The gaps in the spectrograms represent the necessary pulse emissions used for generating the interleaved B-mode images.
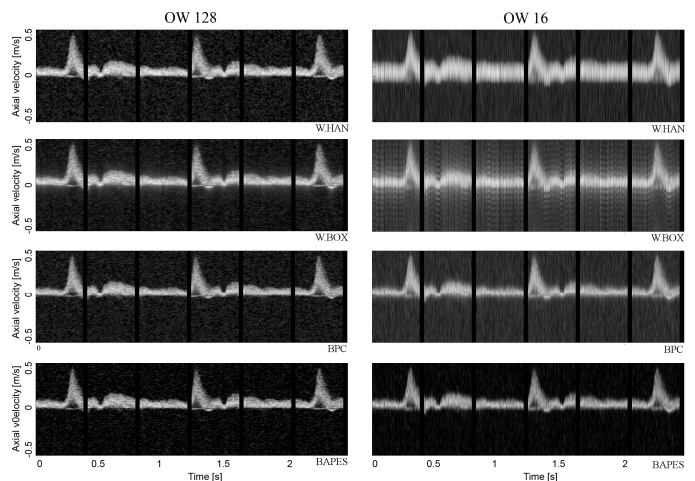

Figure 1: From the same volunteer spectrograms are found using the four methods with OW of 128 and 16.
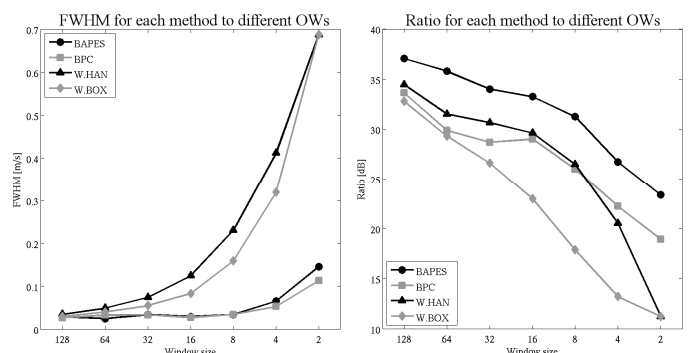

Figure 2: Means for FWHM and ratio are plotted to each window size (OW).

The means of FWHM and ratio for the 280 spectra pooled by method and window are plotted in Fig. 2. From the qualitative evaluation, the total score for each combination of method and OW is found by adding the scores given by nine

radiologists on ten volunteers, thus in the range between 0 and 90. The total score for each combination given in percentage is shown in Fig. 3.
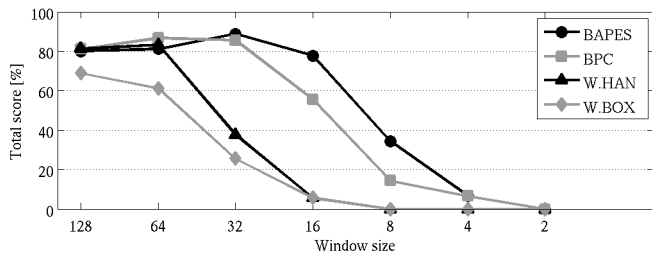

Figure 3: Total sum of scores in percentage for each method to different OWs

The mean of the summed scores for each combination of method and window along with the results of the multiple comparisons are shown in Fig. 4. The sum of scores for each spectrogram is found from the scores given by nine radiologists, and therefore, in the range from 0 to 9. Thus, each combination includes ten summed scores, one for each of the ten spectrograms.
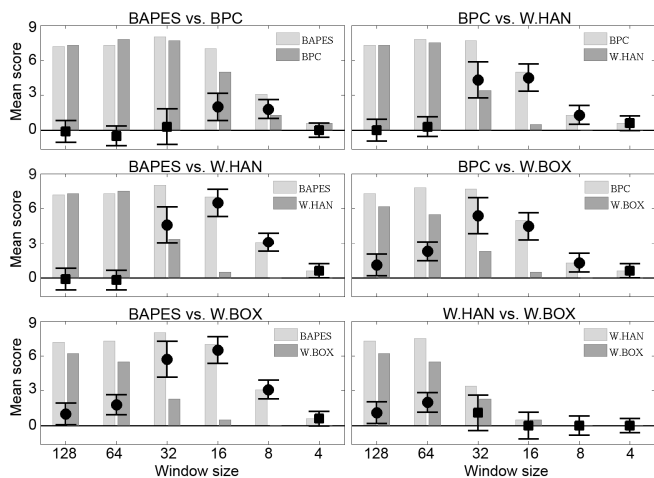

Figure 4: The mean of the summed scores for each method in comparison over different OWs is shown. The 95% confidence interval for the difference in means is illustrated with a bar and marked "■" if insignificant or "●" if significant.

For each method, tests for multiple comparisons were also employed to investigate at which OW the mean score was significantly different from the mean score given at OW 128. W.BOX performed significantly worse when OW was reduced to 64 (*p*=0.03) and W.HAN when OW was reduce to 32 (*p* <0.0001). BPC performed equally well at OW 32 compared to OW 128 (*p*=0.30) while BAPES scored significantly higher at OW 32 compared to OW 128 (*p*=0.02). Only BAPES had equal performance at OW 16 and OW 128 (*p* = 0.55). All four methods decreased in performance at OW 8, 4 and 2 compared to OW 128.

The intra-observer variability for three radiologist evaluating the same 280 spectrograms with more than 14 days apart showed good agreement with an averaged Cohen's

kappa value of 0.79 (94%, κ=0.88; 83%, κ=0.67; 93%, κ=0.81). The inter-observer variability showed moderate agreement using Fleiss' kappa (78%, κ=0.57).

IV. DISCUSSION

In Fig. 1 examples of spectrograms obtained with the four approaches at OW 128 and 16 are shown. Spectral resolution in the spectrogram corresponds to the width of the white curve representing estimated blood velocities, while the difference between the white curve and the darker surroundings corresponds to contrast. At OW 128 the four approaches produced spectrograms of good quality while the estimators differed in performance at OW 16. At OW 16, it is seen that W.BOX estimated with a higher spectral resolution than W.HAN and the adaptive methods with higher spectral resolution than WM. Furthermore, it is seen that the contrast decreased from BAPES to BPC and WM. W.BOX performed poorest among the four methods in terms of contrast and a so-called ringing phenomenon at OW 16 can be seen.

The quantitative tests illustrated in Fig. 2 confirmed the findings in Fig. 1. On average for all OWs above 2, W.BOX performed with better resolution than W.HAN, and W.HAN performed with better contrast than W.BOX. The adaptive methods outperformed WM in terms of resolution at all OWs below 128 and it is seen that BPC had a slightly better resolution than BAPES. In terms of contrast, BAPES was superior compared to both BPC and WM for all OWs. The W.HAN had better contrast than BPC for all OWs above 8.

The overall result of the qualitative evaluation displayed in Fig. 3 shows that the radiologists preferred the adaptive methods over WM, BAPES over BPC and W.HAN over W.BOX. The inter- and intra-observer variability were additionally tested and showed moderate and good agreement, respectively, indicating consistent scores given by each radiologists with a base-line difference among the radiologists. In Fig. 4 the results of the statistical analyses on the mean of the summed scores using Bonferroni adjusted tests for multiple comparisons are displayed. It is seen that at OW 128 and 64, BAPES, BPC and W.HAN performed equally well while W.BOX scored significantly less. When reducing the OW to 32, the adaptive methods BAPES and BPC performed better than W.HAN and W.BOX. However, BAPES was superior to BPC at OW 16 and 8. At OW 16 and 8 the conventional methods were scored useless. When the OW was decreased to 4 and 2 all four methods failed according to the radiologists.

For each method, the scores given at window sizes below OW 128 were compared to the score given at OW 128 to evaluate how robust each method was to reduction of OW. W.BOX scored significantly less at OW 64 while W.HAN at OW 64, BPC at 32 and BAPES at 16 scored as at OW 128.

The radiologists preferred contrast to spectral resolution when comparing scores given for W.BOX and W.HAN at OW 128 and 64 and in most modern scanners the Hanning

weighting scheme is accordingly implemented for spectral blood estimation. However, when comparing BPC and W.HAN the improved resolution in BPC was preferred to the improved contrast in W.HAN. This is seen in Fig. 4 at OW 32 and 16. It indicates that it is not only the raised side lobe level in W.BOX, which reduces the quality of the spectrogram compared to W.HAN but also the accompanying ringing phenomenon.

Moreover, the radiologists preferred BAPES at OW 32 compared to BAPES at OW 128 ($p = 0.01$). The reason could be that the observers were disturbed visually by the high contrast and spectral resolution of the spectrograms obtained with BAPES at OW 128 and somewhat preferred the smoothing of details as presented in spectrograms obtained with BAPES at OW 32.

Fig. 3 shows that even at the presumably best condition, i.e. OW 128, only 80% of the spectrograms were accepted. No dataset was consistently rejected indicating that the ten dataset were of equal quality. However, the ratings of the nine radiologists varied significantly. At OW 128, two of the nine radiologists only accepted 7.5% and 60% of the spectrograms, respectively, while the other seven radiologists on average accepted 90% of the spectrograms (range: 82%-100%). When evaluating the spectrograms acquired on the experimental system, the radiologists use their experience on commercial systems from daily practice as reference. Limitations in the experimental setup are believed to be reflected in the high rejection rate for some of the radiologists. Data was acquired using the experimental scanner RASMUS and the spectrograms were calculated using MATLAB on a 100 CPU Linux cluster. This combination provided a flexible environment for acquiring and processing the data from an experimental point of view, but, compared to commercial scanners, the signal-to-noise ratio of the spectrograms were lower. Furthermore, adjustments of PRF, spectral gain and baseline were not feasible in the evaluation setup.

The spectral ultrasound examination of today can potentially be improved if the conventional WM for blood velocity estimation used in commercial scanners was to be replaced with one of adaptive spectral estimators BPC or BAPES. The reduction in OW could potentially be used to increase the temporal resolution of the spectrogram or the frame rate of the accompanying B-mode images.

The adaptive methods are more computationally demanding than the conventional WM, and BAPES is more computationally demanding than BPC. However, when reducing the OW, the computational costs decrease as well. Therefore, the BAPES method implemented with an OW 16 or the BPC method with an OW 32 may be as tractable as WM with a much longer OW. It should therefore be possible to implement the methods into commercial scanners and convert the methods to an actual real-time modality.

V. CONCLUSION

In this paper the performances of two adaptive spectral estimators BPC and BAPES were investigated *in-vivo*. It was shown that the adaptive methods were superior to the conventional WM and that BAPES was superior to BPC. The quantitative tests showed that BAPES and BPC had improved spectral resolution and BAPES had improved spectral contrast compared to WM. According to the scores given by nine radiologists, OW can be reduced to 32 when using BPC, and 16 when using BAPES method for estimating spectrogram without losing performance. The results indicate that the adaptive methods BPC and BAPES potentially can bring improvements to spectral blood estimation as an increase of the temporal resolution of the spectrogram or as an increase of the frame rate for the interleaved B-mode images. Further studies are needed to evaluate the adaptive spectral estimators in various *in-vivo* settings with different flow profiles, vessel geometries and scan depths using different transducers.

REFERENCES

[1] P.D.Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Trans Audio Electroacoust., vol. AU-15, pp. 70-73, 1967.

[2] J.B.Kruskal, P.A.Newman, L.G.Sammons, and R.A.Kane, "Optimizing Doppler and color flow US: Application to hepatic sonography," Radiographics, vol. 24, no. 3, pp. 657-675, 2004.

[3] A.J.S.Coats, C.Murphy, J.Conway, and P.Sleight, "Validation of the beat to beat measurement of blood velocity in the human ascending aorta by a new high temporal resolution Doppler ultrasound spectral analyser," Br. Heart J, vol. 68, no. 8, pp. 223-229, Aug.1992.

[4] P.Stoica and R.Moses, Introduction to spectral analysis. Upper Saddle River, N.J: Prentice Hall, 1997.

[5] F.Gran, A.Jakobsson, and J.A.Jensen, "Adaptive spectral Doppler estimation," IEEE Trans Ultrason Ferroelec Freq Contr, vol. 56, no. 4, pp. 700-714, 2009.

[6] J.A.Jensen, O.Holm, L.J.Jensen, H.Bendsen, S.Nikolov, B.G.Tomov, P.Munk, M.Hansen, K.Salomonsen, J.Hansen, K.Gormsen, H.M.Pedersen, and K.L.Gammelmark, "Ultrasound research scanner for real-time synthetic aperture image acquisition," IEEE Trans Ultrason Ferroelec Freq Contr, vol. 52, no. 5, pp. 881-891, 2005.

[7] D.G.Altman, Practical statistics for medical research. London: Chapmann & Hall/CRC, 1991.

[8] J.R.Landis and G.G.Koch, "The measurement of observer agreement for categorical data," Biometrics, vol. 33, pp. 159-174, 1977.

[9] J.L.Fleiss, "Measuring nominal scale agreement among many raters," Psychol. Bull., vol. 76, no. 5, pp. 378-382, 1971.