Aalto University

OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Bäckström, T.

Overlap-add Windows with Maximum Energy Concentration for Speech and Audio Processing

Published in:
ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

DOI:
10.1109/ICASSP.2019.8683577

Published: 01/05/2019

Document Version
Peer reviewed version

Please cite the original version:
Bäckström, T. (2019). Overlap-add Windows with Maximum Energy Concentration for Speech and Audio Processing. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 491-495). [8683577] (Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing). IEEE. https://doi.org/10.1109/ICASSP.2019.8683577

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

This is the accepted version of the original article published by IEEE.

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

OVERLAP-ADD WINDOWS WITH MAXIMUM ENERGY CONCENTRATION FOR SPEECH AND AUDIO PROCESSING

Tom Bäckström

Aalto University, Department of Signal Processing and Acoustics, Espoo, Finland

ABSTRACT

Processing of speech and audio signals with time-frequency representations require windowing methods which allow perfect reconstruction of the original signal and where processing artifacts have a predictable behavior. The most common approach for this purpose is overlap-add windowing, where signal segments are windowed before and after processing. Commonly used windows include the half-sine and a Kaiser-Bessel derived window. The latter is an approximation of the discrete prolate spherical sequence, and thus a maximum energy concentration window, adapted for overlap-add. We demonstrate that performance can be improved by including the overlap-add structure as a constraint in optimization of the maximum energy concentration criteria. The same approach can be used to find further special cases such as optimal low-overlap windows. Our experiments demonstrate that the proposed windows provide notable improvements in terms of reduction in side-lobe magnitude.

Index Terms— time-frequency processing, windowing, discrete prolate spherical sequences

1. INTRODUCTION

Speech and audio signals are slowly time-varying in character, such that it is beneficial to analyze and process them in short segments. When the segment length is chosen appropriately, we can treat the signal as a stationary process within the segment such that statistical modeling becomes efficient. Many applications then use time-frequency transforms on the segments such as the short-time Fourier transform or the modified discrete cosine transform, for the benefit of statistical and perceptual efficiency [1–4].

Segmenting a signal is a windowing problem, where the segment is extracted by multiplying with a windowing function, which is non-zero in a limited range. In analysis applications, signal processing has a long history in the design of such windowing functions and its theory is presented in every basic book of signal processing, e.g. [5]. The principal objective of windowing in analysis applications is to minimize the detrimental effect of windowing on the signal statistics. In processing applications, however, we also need to consider the effect of windowing on the reconstruction process.

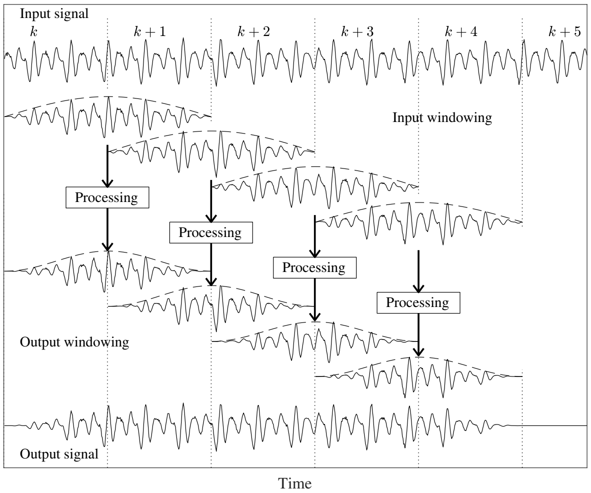

Fig. 1. Illustration of input and output windowing with overlap-add synthesis in a speech processing application.

A widely used approach in time-frequency processing of signals is known as overlap-add, where the input signal is windowed into overlapping segments, and after processing, the segments are windowed a second time before adding them together [4, 6, 7] (see Fig. 1). By a careful choice of windowing functions, we can ensure that, in the absence of modifications to the windowed signal, the original signal can be reconstructed from the windowed segments. This is known as the *perfect reconstruction* property.

Windowing is often discussed in combination with time-frequency transforms, whence the combination is known as a filterbank [8]. A particular type of filterbanks are those which, in addition to perfect reconstruction, also provide critical sampling. The most commonly used critically sampled filterbank in audio processing is the modified discrete cosine transform [9–11], which can also be applied in a bit-exact manner [12]. Typically, such applications use the half-sine or a Kaiser-Bessel derived (KBD) window, which are some of the few windows applicable in overlap-add. A radically different approach is commonly used in speech coding with code-excited linear prediction (CELP), where temporal corre-

This project was supported by the Academy of Finland research project No 312490.

lation is explicitly modeled by a linear predictor, such that the predictor residual can be windowed without overlaps [3, 13].

The performance of windows which are suitable for overlap-add have however not received the same rigorous attention as the classical windowing methods. This paper presents methods for representing the symmetries required by overlap-add as constraints such that the window performance can be optimized. Specifically, we will use the maximum energy concentration criteria [14], familiar from Slepian or discrete prolate spherical sequence (DPSS) -windows, to obtain optimal windows for overlap-add.

2. OVERLAP-ADD WINDOWING

The objectives, when applying windowing in processing applications, are two-fold: 1. In the absence of any modifications, we require that the original signal can be reconstructed perfectly. 2. When the windowed signal is modified, then the energy expectation of the modification (or error), in the output signal, should be uniform over time.

Let x_k be our input signal which we want to segment into overlapping windows. Window h of the signal is then

$$y_{k,h} = w_{k-Lh/2} x_k, \tag{1}$$

where w_k is the windowing function of length L for which

$$\begin{cases} w_k > 0, & \text{when } k \in [1, L] \\ w_k = 0, & \text{when } k \leq 0 \text{ or } k > L. \end{cases} \tag{2}$$

We can then apply some processing on the windows $y_{k,h}$ such that the modified signal is $\hat{y}_{k,h}$.

To reconstruct the signal, we apply windowing again by multiplying with the windowing function and add the windows together, such that the modified output signal is

$$\hat{x}_k := \sum_h w_{k-Lh/2} \hat{y}_{k,h}. \tag{3}$$

It is important to observe that the window is applied twice, once on the input signal and a second time after processing on the modified output window. Only after applying the window twice can we add the segments together to obtain the resynthesised signal.

It is well-known and we can readily see that both the requirement of perfect reconstruction and uniform error energy is ensured when the windowing function satisfies the Princen-Bradley criteria [2, 3]

$$w_{k+L/2}^2 + w_k^2 = 1, \qquad \text{for } k \in [1,\, L/2]. \tag{4}$$

Figure 2 illustrates a typical windowing function which satisfies the Princen-Bradley criteria and Figure 1 illustrates the effect of overlap-add windowing on a speech signal.

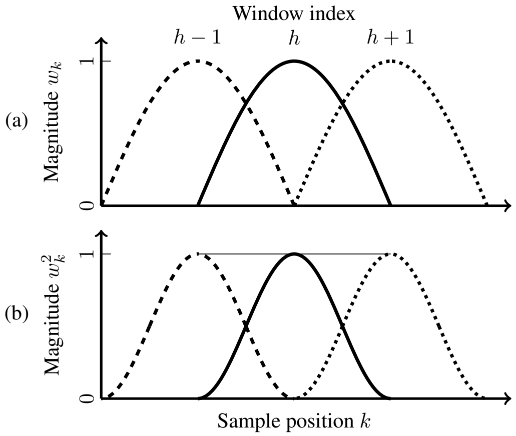

Fig. 2. (a) Typical input windowing functions of subsequent frames and (b) the corresponding squared windowing functions. The thin line in (b) demonstrates the region where the windows add up to unity as required by Eq. 4.

3. CONSTRAINED MAXIMIZATION OF ENERGY CONCENTRATION

Windowing in the time-domain corresponds to convolution in the frequency-domain. To minimize frequency-domain distortion, we therefore require that energy of the windowing function in the frequency-domain is maximally concentrated. The concentration of energy can be evaluated by the ratio of energy in the pass-band versus total energy

$$\tau = \frac{\int_{-\delta}^{\delta} |W(f)|^2 df}{\int_{-\infty}^{\infty} |W(f)|^2 df}, \tag{5}$$

where $W(f)$ is the spectrum of the windowing function and δ is the bandwidth of the pass-band. For discrete, finite length windowing functions $\mathbf{w} = [w_1, \dots, w_L]^T$ it can be shown that the above ratio is equivalent with

$$\tau = \frac{\mathbf{w}^T \mathbf{T} \mathbf{w}}{\|\mathbf{w}\|^2}, \tag{6}$$

where $\mathbf{T} \in \mathbb{R}^{L \times L}$ is a symmetric Toeplitz matrix with elements

$$\mathbf{T}_{k-h} = \frac{L \sin\left(\frac{\pi}{L}\alpha(k-h)\right)}{(k-h)}, \tag{7}$$

where $\alpha \in (0, 1)$ defines the width of the main lobe. Clearly the maximum of τ is then the eigenvector of $\mathbf{T}$ corresponding to the largest eigenvalue and we can equivalently define

$$\max \mathbf{w}^T \mathbf{T} \mathbf{w} \text{ such that } \|w\|^2 = 1. \tag{8}$$

The eigenvectors of $\mathbf{T}$ are known as discrete prolate spherical sequences (DPSS) and the corresponding windowing functions are known correspondingly as DPSS or Slepian windows [14–16].

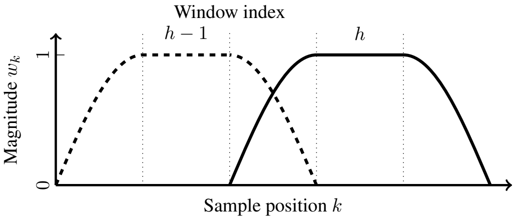

Fig. 3. To obtain a low overlap between windows, we can constrain a number of samples in the middle of the window to have unit magnitude. Thin dotted vertical lines indicate the borders between the flat tops and overlap areas.

The main objective of this paper is to design windowing functions which fulfill those symmetries required by overlap-add processing, while simultaneously optimizing the above spectral characteristics. First, the Princen-Bradley conditions of Eq. 4 can then be written as

$$\mathbf{w}^T\mathbf{P}_k\mathbf{w} = 1, \tag{9}$$

where $\mathbf{P}_k$ is diagonal with diagonal entries $[\mathbf{P}_k]_{h,h} = \delta_{k-h} + \delta_{k+L/2-h}$. In other words, $\mathbf{P}_k$ has two non-zero entries on the diagonal which pick out the kth and $(k+L/2)$th samples of the windowing vector $\mathbf{w}$. Consequently, the matrices $\mathbf{P}_k$ are positive semi-definite. Observe that the constraints Eq. 9 is similar to the constraint in Eq. 8 but more strict. We can therefore define a new optimization problem, using the constraints of Eq. 9 and the objective function of Eq. 8 as

$$\max \mathbf{w}^T\mathbf{T}\mathbf{w} \text{ such that } \mathbf{w}^T\mathbf{P}_k\mathbf{w} = 1 \text{ for } k \in [1,\ L/2]. \tag{10}$$

This is a quadratically constrained quadratic programming (QCQP) problem, which is known to be convex if the matrices $\mathbf{T}$ and $\mathbf{P}_k$ are positive definite. We can therefore use numerical optimization based interior-point methods to find the optimal solution.

4. LOW-OVERLAP WINDOWS

In some applications it is desirable to limit the overlap length between windows [17]. The conventional approach in designing windows of length L with overlap T, is to choose a windowing function of length $2T$ and extend it by a vector of $L-2T$ ones in the middle, such that the desired length is achieved (see Fig. 3). This heuristic method can now be amended using the optimization presented above.

Specifically, we can define new constraints as

$$\begin{cases} w_{k+L-T}^2 + w_k^2 &= 1, \quad \text{for } k \in [1,\ T]. \\ \qquad\qquad w_k &= 1, \quad \text{for } k \in (T, L-T). \end{cases} \tag{11}$$

Substituting these quadratic and linear constraints into the optimization problem of Eq. 10 yields a low-overlap window which has maximal energy concentration.

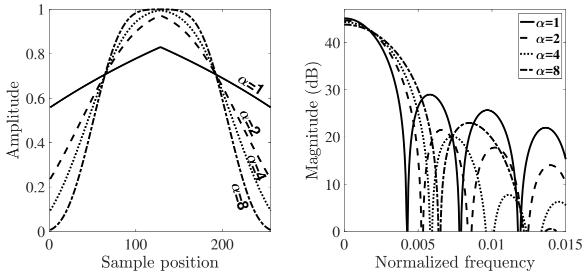

Fig. 4. Illustration of the shapes of proposed window with different values of α and a window length of $L = 256$.

5. EVALUATION

The most commonly used overlap-add windows include the half-sine and a Kaiser-Bessel derived (KBD) window. The half-sine window is defined as

$$w_{\sin,k} = \sin\left(\frac{\pi\left(k-\frac{1}{2}\right)}{L}\right), \text{ for } k \in [1, L]. \tag{12}$$

The KBD window is based on the Kaiser-Bessel window, defined as

$$u_k = I_0\left[\alpha\sqrt{1-\left(\frac{2(k-\frac{1}{2})}{L-1}-1\right)^2}\,\right], \text{ for } k \in [1, L], \tag{13}$$

where $\alpha > 0$ specifies the width of the main-lobe. The KBD window is then defined as

$$w_{KBD,k} = \sqrt{\frac{\sum_{h=1}^{k} u_k}{\sum_{h=1}^{L} u_k}}, \text{ for } k \in [1, L]. \tag{14}$$

In other words, the KBD takes the cumulative sum of the Kaiser-Bessel window, normalizes it by the sum and then takes a square root to satisfy Princen-Bradley.

We generated the proposed DPSS based overlap-add windows (OLA-DPSS) by using the interior-point algorithm of the Optimization toolbox in Matlab2018a. Fig. 4 demonstrates the obtained window shapes for different values of the parameter α. As an informal observation, we did not have any problems with convergence and the running times were only some seconds. Since windowing functions are usually determined off-line, we conclude that computational capacity is not an issue in calculation of OLA-DPSS windows.

Figure 5 illustrates the half-sine, KBD and the proposed windows and their spectral responses. Note that we have here manually tuned the pass-band bandwidth α's in Eqs. 13 and Eq. 7 such that the main-lobe widths match that of the half-sine window. This choice allows fair comparison of the side-lobe magnitudes.

We observe that the KBD window is in shape very similar to the half-sine, and their spectral responses differ only for

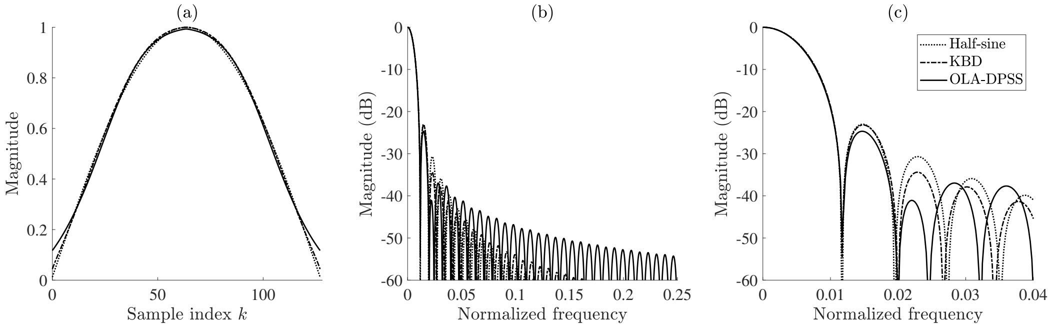

Fig. 5. Illustration of (a) the half-sine, KBD and proposed OLA-DPSS windows, (b) their spectral responses and (c) responses focused on the central region. Window length is $L = 128$ and KBD has $\alpha = 4.25$ and OLA-DPSS has $\alpha = 2.75$.

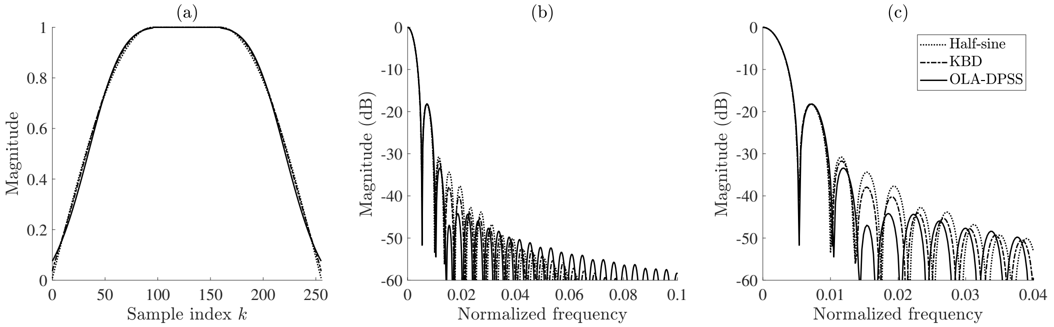

Fig. 6. Illustration of low-overlap versions of (a) the half-sine, KBD and proposed OLA-DPSS windows, (b) their spectral responses and (c) responses focused on the central region. Window length is $L = 256$, overlap length is $T = 64$, KBD has $\alpha = 4.25$ and OLA-DPSS has $\alpha = 5$.

the second side-lobe and higher. The shape of the proposed OLA-DPSS window, however has slightly higher tails near the ends of the window. Moreover, the spectral response of the OLA-DPSS has an approximately 2 dB benefit for the first side-lobe. The energy concentration ratios following Eq. 6, for the half-sine, KBD and OLA-DPSS windows are 16.6559, 16.6582 and 16.6624 dB (parameters as in Fig. 5). In other words, by using OLA-DPSS, we obtain 0.0065 dB and 0.0041 dB improvements in energy concentration in comparison to the half-sine and KBD windows respectively.

Figure 6 illustrates low-delay versions of the half-sine, KBD and the proposed windows and their spectral responses. Here we find differences only from the second side-lobe, where the OLA-DPSS is about 3 dB better than the half-sine and 2 dB better than KBD. The corresponding energy concentration ratios are 19.6191, 19.6182, 19.6193 dB indicating that again the OLA-DPSS is the best (by design) but the difference to the others is marginal.

6. CONCLUSIONS

Design of windowing functions has a long tradition in signal analysis. In processing of speech and audio signals, we however require that reconstruction of signals is possible. The conventional approach is to use a method known as overlap-add, where subsequent windows are overlapped such that their sum recovers the original signal. This places constraints on the window design which has not been adequately taken into account in previous studies.

Slepian windowing functions based on discrete prolate spherical sequences (DPSS) are optimal in terms of energy concentration, whereby we propose to apply the same objective function but with constraints that satisfy the symmetries required by overlap-add. The optimization problem is a quadratically constrained quadratic programming problem, whose solution has become feasible with modern optimization toolboxes. Since windowing functions are usually determined off-line, computational complexity is not an issue.

The presented evaluations confirm that the proposed overlap-add DPSS or OLA-DPSS windows are efficient in energy concentration as desired and the proposed window is better than the conventional windows in all comparisons presented. Since the proposed overlap-add window surpasses the performance of conventional windows in all aspects, indeed it is the optimal window for this application, OLA-DPSS should be the preferred choice in speech and audio processing applications.

7. REFERENCES

[1] J Benesty, M Sondhi, and Y Huang, *Springer Handbook of Speech Processing*, Springer, 2008.

[2] M Bosi and R E Goldberg, *Introduction to Digital Audio Coding and Standards*, Kluwer Academic Publishers, 2003.

[3] T Bäckström, *Speech Coding with Code-Excited Linear Prediction*, Springer, 2017.

[4] J Vilkamo and T Bäckström, "Time-frequency processing: Methods and tools," in *Parametric Time-Frequency Domain Spatial Audio*, V Pulkki, S Delikaris-Manias, and A Politis, Eds., pp. 3–24. Wiley, 2017.

[5] S K Mitra, *Digital signal processing: a computer-based approach*, McGraw-Hill, 1998.

[6] F J Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, 1978.

[7] AH Nuttall, "Spectral analysis by means of overlapped fast Fourier transform processing of windowed data," Tech. Rep., NUSC Tech. Rep, 4169, 1971.

[8] B Boashash, *Time-frequency signal analysis and processing: a comprehensive reference*, Academic Press, 2015.

[9] B Edler, "Codierung von audiosignalen mit Überlappender transformation und adaptiven fensterfunktionen," *Frequenz*, vol. 43, no. 9, pp. 252–256, 1989.

[10] H S Malvar, "Lapped transforms for efficient transform/subband coding," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 6, pp. 969–978, 1990.

[11] H S Malvar, *Signal processing with lapped transforms*, Artech House, Inc., 1992.

[12] R Geiger, T Sporer, J Koller, and K Brandenburg, "Audio coding based on integer transforms," in *Audio Engineering Society Convention 111*. Audio Engineering Society, 2001.

[13] T Bäckström, "Comparison of windowing in speech and audio coding," in *Proc. WASPAA*, Oct. 2013.

[14] L C Barbosa, "A maximum-energy-concentration spectral window," *IBM journal of research and development*, vol. 30, no. 3, pp. 321–325, 1986.

[15] D Slepian and H O Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—i," *Bell System Technical Journal*, vol. 40, no. 1, pp. 43–63, 1961.

[16] F J Simons, "Slepian functions and their use in signal estimation and spectral analysis," in *Handbook of geomathematics*, pp. 891–923. Springer, 2010.

[17] G Fuchs, C R Helmrich, G Marković, M Neusinger, E Ravelli, and T Moriya, "Low delay LPC and MDCT-based audio coding in the EVS codec," in *Proc. ICASSP*, 2015, pp. 5723–5727.