Spectral Windowing for Enhanced Temporal Noise Shaping Analysis in Transform Audio Codecs

Richard Füg
International Audio Laboratories Erlangen

***Abstract*—Temporal Noise Shaping (TNS) is a method employed in many transform audio codecs. It works by performing open-loop linear predictive coding (LPC) in the frequency domain. In state-of-the-art codecs the Autocorrelation Function (ACF) required for the determination of the LPC coefficients during TNS anlysis is calculated on a spectral region cut out with a de-facto rectangular window. In this work a detailed mathematical description of the TNS analysis is done to show that rectangular spectral windowing leads to a suboptimal temporal resolution of TNS. Furthermore a new windowing scheme is proposed that allows for a good tradeoff between temporal and spectral resolution for TNS. The effect on the perceptual quality for different variants of this scheme is evaluated with a listening test in an experimental coding environment.**

***Index Terms*—audio coding, modified discrete cosine transform (MDCT), temporal noise shaping (TNS), linear prediction in spectral domain, frequency-domain linear prediction (FDLP)**

I. Introduction

Widely used and state-of-the-art audio codecs like AAC [1], USAC [2] and MPEG-H [3] make use of transform audio coding using the MDCT (Modified Discrete Cosine Transform) [4] [5]. While this allows for fine control of the spectral distribution of the quantization noise, control of the temporal distribution is not straightforward, often resulting in the so-called pre-echo artifact. It occurs when a transient signal is coded with a transform audio codec and the resulting quantization noise is spread out over the entire transform window, whereas the auditory masking effect is only strong during a short duration around the transient. [6]

To tackle this issue, several methods were introduced, that require a significantly higher number of bits for transmission (pre-echo control by spending more bits on transient segments), an additional filter bank (gain control [7], HREP [8]) or additional delay (window length switching [9]). In contrast to these methods, the Temporal Noise Shaping (TNS) method requires comparatively low bitrate overhead, works in the coding transform domain, does not require any additional delay and can be combined with the other methods easily [10]. Therefore TNS is a part of all codecs mentioned above.

The idea of TNS is to perform open-loop Linear Predictive Coding (LPC) as known from time-domain coding in the frequency domain. This is shown for a simplified transform codec in Figure 1. Open-loop LPC in the time domain is known to result in a quantized signal where the quantization noise is spectrally shaped similar to the original signal. TNS – being the dual of this – leads to a quantized signal where the quantization noise is shaped temporally similarly to the original signal.

More specifically, the linear predictive filtering on the spectrum leads to a temporal flattening of the corresponding time-domain signal in the encoder before quantization. The quantization noise added to the spectrum afterwards can also be considered rather temporally flat (which without TNS usually results in the pre-echo problem). At the decoder, the inverse linear prediction filter then restores the original temporal envelope of the signal while also shaping the quantization noise accordingly.

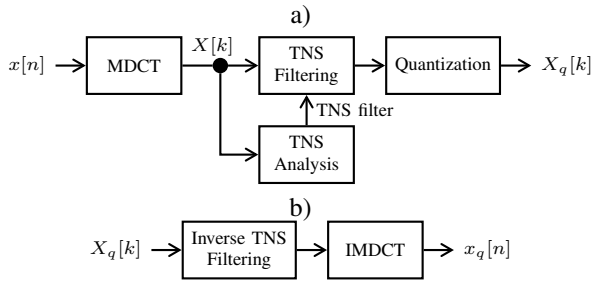

Fig. 1: Simplified transform audio encoder (a) and decoder (b) using TNS. Transmission of side information and TNS filter is not shown.

In order to obtain the linear prediction filter for TNS – a process referred to as TNS analysis in the following – the spectral Autocorrelation Function (ACF) is calculated on the spectrum first. Then the Levinson-Durbin algorithm is used to solve the normal equations which results in the coefficients of the linear prediction filter. [11]

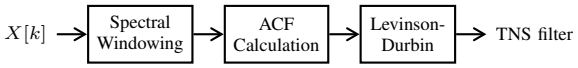

Fig. 2: Block diagram showing the basic steps involved in TNS analysis.

As TNS is applied on the MDCT spectrum it is straightforward to use different filters on different frequency ranges and thereby account for different temporal behavior of the signal in these ranges. To create these different filters, the spectral ACF has to be calculated on these frequency ranges. In all TNS implementations known to the author this is done by (implicit) rectangular windowing of the spectrum for these ranges. Figure 2 shows a block diagram for the steps involved in TNS analysis. Just like rectangular windowing in time domain can be a bad choice for analysis of the spectrum due

to strong spectral leakage, the dual holds true for rectangular windowing in frequency domain and strong temporal leakage.

While TNS works very well for a lot of transient signals, the rectangular window in the frequency domain during TNS analysis strongly reduces the temporal noise shaping effect for very sharp transient signals and therefore might still lead to perceivable pre-echo artifacts.

Intuition from spectral analysis [12] and linear predictive coding in time-domain as well as results from Frequency Domain Linear Prediction (FDLP) [13] suggest that using non-rectangular windows for the TNS analysis can lead to enhanced temporal resolution and better shaping.

Therefore, in this work we will first lay out the theoretical foundation that allows for the explanation of the effect of the spectral window on the TNS analysis in the MDCT domain using the Generalized Discrete Fourier Transform (GDFT). Then we will analyze the deficiencies of the spectral rectangular window followed by a proposal for a windowing scheme with more suitable properties for TNS. Finally, this new windowing scheme is evaluated in a listening test.

II. GDFT AND ITS RELATION TO MDCT

The relation between MDCT and GDFT, which is sometimes also referred to as Shifted Discrete Fourier Transform (SDFT), has been explored in great detail and is therefore only summarized here [14] [15] [16]. Due to the close relationship between GDFT and all Discrete Cosine/Sine Transforms (DCTs/DSTs) [17], all theory laid out for the MDCT in this work can also be easily applied to these transforms as well.

For a time-domain input signal $x[n]$ with $n = 0 \dots M-1$ the GDFT

$$X[k] = \mathcal{G}_{a,b}\left\{x\left[n\right]\right\} = \sum_{n=0}^{M-1} x\left[n\right] \cdot e^{-j\frac{2\pi}{M}\cdot(k+a)\cdot(n+b)} \quad (1)$$

and the inverse GDFT are defined as

$$x[n] = \mathcal{G}_{a,b}^{-1}\left\{X\left[k\right]\right\} = \frac{1}{M}\sum_{k=0}^{M-1} X\left[k\right] \cdot e^{j\frac{2\pi}{M}\cdot(k+a)\cdot(n+b)} \quad (2)$$

where $k = 0 \dots M-1$ with M being an even number. a and b are arbitrary real-valued numbers [17]. For the sake of a concise notation we assume $x[n]$ to be temporally windowed.

For the special case where $a = 1/2$ and $b = 1/2 + N/2$ where $N = M/2$, the first N values of $\mathfrak{Re}\left\{X[k]\right\}$ are identical to the corresponding MDCT spectrum up to a multiplicative factor. Alternatively, the first N values of $X[k]$ are identical to the MDCT if $x[n]$ follows the Time Domain Aliasing (TDA) symmetries of the MDCT.

We can also introduce the analytic spectrum [17] corresponding to $X[k] = \mathcal{G}_{\frac{1}{2},\frac{N+1}{2}}\left\{x\left[n\right]\right\}$ as

$$\mathring{X}[k] = \begin{cases} X[k] & \text{for } 0 \le k < M/2 \\ 0 & \text{for } M/2 \le k < M \end{cases} \quad (3)$$

and the corresponding time-domain signal as

$$\mathring{x}[n] = \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}}\left\{\mathring{X}[k]\right\}. \quad (4)$$

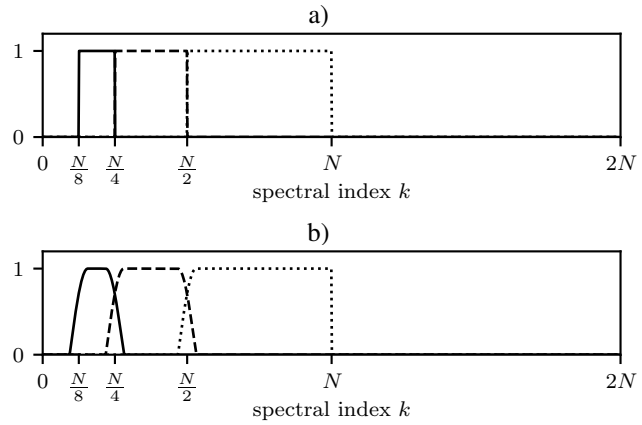

Fig. 3: Example for spectral windowing styles before ACF calculation for TNS on three frequency ranges using rectangular windows (a) and overlapping windows (b). The frequency bin ranges are $[N/8, N/4[$ (–), $[N/4, N/2[$ (- -) and $[N/2, N[$ ($\cdots$).

Trivially, the relations of MDCT and GDFT also hold true for the analytic spectrum.

For the special case of GDFT considered here, we can introduce the convolution-multiplication properties

$$M \cdot \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}}\left\{X[k]\right\} \cdot \mathcal{G}^{-1}_{0,\frac{N+1}{2}}\left\{Y[k]\right\} = \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}}\left\{X[k] \overset{s}{\otimes} Y[k]\right\} \quad (5)$$

and

$$\mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}}\left\{X[k]\cdot Y[k]\right\} = \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}}\left\{X[k]\right\} \overset{s}{\otimes} \mathcal{G}^{-1}_{\frac{1}{2},0}\left\{Y[k]\right\} \quad (6)$$

where $X[k]$ and $Y[k]$ are spectra of length M and $\overset{s}{\otimes}$ represents skew-circular convolution [18].

Similarly for calculating the skew-circular spectral ACF we introduce

$$\mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}}\left\{X[k]\right\} \cdot \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}}\left\{X[k]\right\}^* = \mathcal{G}^{-1}_{0,\frac{N+1}{2}}\left\{X[k] \overset{s}{\circledast} X[k]\right\} \quad (7)$$

where $\overset{s}{\circledast}$ denotes the skew-circular correlation [17]. These properties can be derived in a straightforward way from the properties of the GDFT as given in [17] [18] [19].

Using Equations 7 and 4 we can express the spectral ACF of the analytic spectrum to

$$R_{\mathring{X},\mathring{X}}[k] = \mathring{X}[k] \overset{s}{\circledast} \mathring{X}[k] = \mathcal{G}_{0,\frac{N+1}{2}}\left\{\left|\mathring{x}[n]\right|^2\right\} \quad (8)$$

where $\left|\mathring{x}[n]\right|^2$ is also known as the squared Hilbert envelope corresponding to the signal $x[n]$. This expression is the $\mathcal{G}_{\frac{1}{2},\frac{N+1}{2}}$ version of the motivation for TNS given in [10] for the continuous Fourier transform.

If $\mathring{X}[k]$ is identical to an MDCT spectrum $Z[k]$, then $R_{\mathring{X},\mathring{X}}$ is identical to the ACF that is calculated on $Z[k]$ using the well-known biased ACF formula assuming zeros outside of the borders of $Z[k]$.

III. SPECTRAL WINDOWING IN TNS ANALYSIS

Using the results of the previous section we can now analyze the influence of spectral windowing during TNS analysis on

the spectral ACF and the resulting TNS filter – in general and for the state-of-the-art windowing scheme.

Henceforth we will consider $\mathring{X}[k]$ to be a $\mathcal{G}_{\frac{1}{2},\frac{N+1}{2}}$ spectrum of length M, where the first N spectral values are the values of the MDCT spectrum to be analyzed and all other values are zero. We define the analytic spectrum

$$\mathring{X}_w[k] = \mathring{X}[k] \cdot W[k] \tag{9}$$

which is the analytic spectrum $\mathring{X}[k]$ windowed with the arbitrary spectral window $W[k]$ of length M. Using equations 4 and 6 we calculate

$$\begin{aligned}\mathring{x}_w[n] =& \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}} \left\{ \mathring{X}_w[k] \right\} \\ =& \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}} \left\{ \mathring{X}[k] \cdot W[k] \right\} \\ =& \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}} \left\{ \mathring{X}[k] \right\} \overset{s}{\otimes} \mathcal{G}^{-1}_{\frac{1}{2},0} \left\{ W[k] \right\} \\ =& \mathring{x}[n] \overset{s}{\otimes} w[n]\end{aligned} \tag{10}$$

for the analytic signal. By inspection of Equation 10 the temporal response of this window can be calculated via

$$w[n] = \mathcal{G}^{-1}_{\frac{1}{2},0} \left\{ W[k] \right\}. \tag{11}$$

The spectral ACF for the windowed analytic spectrum results in

$$\begin{aligned}R_{\mathring{X}_w,\mathring{X}_w}[k] =& \mathring{X}_w[k] \overset{s}{\circledast} \mathring{X}_w[k] \\ =& \mathcal{G}^{-1}_{0,\frac{N+1}{2}} \left\{ \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}} \left\{ \mathring{X}_w[k] \right\} \cdot \mathcal{G}^{-1}_{\frac{1}{2},\frac{N+1}{2}} \left\{ \mathring{X}_w[k] \right\}^* \right\} \\ =& \mathcal{G}^{-1}_{0,\frac{N+1}{2}} \left\{ |\mathring{x}_w[n]|^2 \right\}.\end{aligned} \tag{12}$$

This expression shows that the application of the spectral window before calculation of the spectral ACF results in a spectral ACF corresponding to the Hilbert envelope $|\mathring{x}_w[n]|$ of an analytic signal that was convolved with the temporal response of the spectral window.

During the TNS analysis this spectral ACF is used in the Levinson-Durbin algorithm to obtain the TNS filter coefficients $A[k]$ which we consider to be zero-padded to length M. The TNS envelope – the envelope that TNS will shape the quantization noise to – can be calculated to

$$e[n] = \frac{1}{M \cdot \left| \mathcal{G}^{-1}_{0,\frac{N+1}{2}} \left\{ A[k] \right\} \right|}. \tag{13}$$

The spectral rectangular windowing inherently employed in state-of-the-art audio codecs is shown in Figure 3a. The window function for the respective frequency bin range $[v, w[$ is given as

$$W_{\text{rect},v,w}[k] = \begin{cases} 1 & \text{for } v \leq k < w \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where $v < w < N$. Figure 4 shows the temporal response of such a spectral rectangular window, which shows a rather flat behavior. Consequently the convolution with the analytic signal will result in a Hilbert envelope with a lot of temporal leakage as shown in Figures 5b and 6b. As this envelope is directly connected to the spectral ACF and therefore the basis for the calculation of the TNS filter, the temporal leakage decreases the shaping quality of the filter. This can be observed in Figures 5c and 6c.

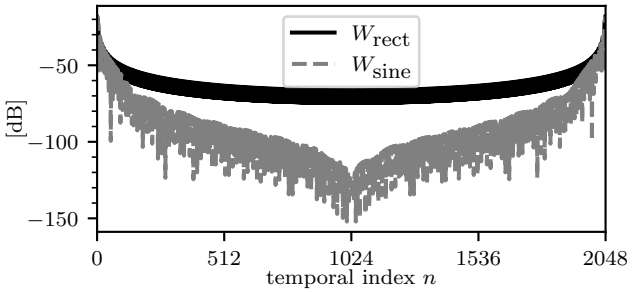

Fig. 4: Temporal response magnitude for exemplary rectangular window $W_{\text{rect},N/4,N/2}[k]$ and window with sine overlap $W_{\text{sine},N/4,N/2}[k]$ where overlap lengths $l = r = 16$ and $M = 2N = 2048$.

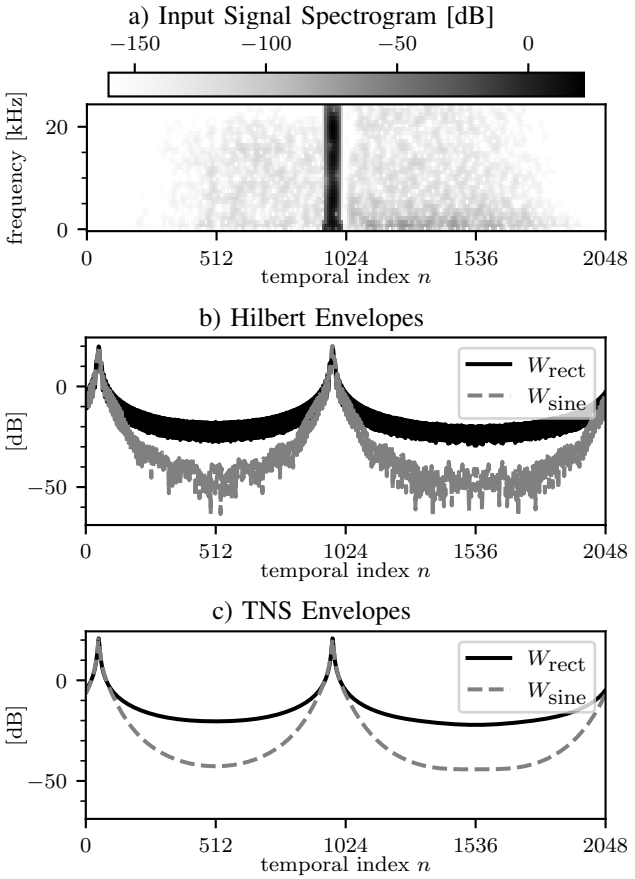

Fig. 5: Spectrogram of an impulse-like input signal (a) of length $M = 2N = 2048$ to the MDCT, resulting Hilbert envelopes (b) and TNS envelopes for filter order 8 (c) for an exemplary rectangular window and a window with sine overlap with the same specifications as in Figure 4. The Hilbert and TNS envelopes were normalized to an RMS of 1. Note that the mirrored impulse at the very left side in (b) and (c) is due to time domain aliasing of the MDCT.

IV. New Scheme for Spectral Windowing

In order to overcome the temporal leakage issue with the spectral rectangular window, a new scheme is introduced in this section. It is inspired by codecs using adaptive LPC in time domain, where overlapping windows are used for the calculation of the temporal ACF in order to overcome the influence of rectangular windowing. Therefore a new windowing scheme for TNS analysis as shown in Figure 3b is proposed. In contrast to using non-overlapping rectangular windows, overlapping windows with a smooth roll-off for the TNS frequency ranges are used. It has to be emphasized that only the TNS analysis is affected by the windowing and the actual filtering is not changed.

A window using a sine shape for the overlapping region can be defined as

$$W_{\text{sine},v,w}[k] = \begin{cases} s_l[k - v - \frac{l}{2}] & \text{for } v - \frac{l}{2} \leq k < v + \frac{l}{2} \\ 1 & \text{for } v + \frac{l}{2} \leq k < w - \frac{r}{2} \\ s_r[k - w + \frac{r}{2}] & \text{for } w - \frac{r}{2} \leq k < w + \frac{r}{2} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where l and r are even integer numbers specifying the left and right overlap length respectively. Furthermore

$$s_L[k] = \sin\left(\frac{k + 0.5}{L}\pi\right) \text{ with } k = 0 \ldots L - 1 \tag{16}$$

is the sine window.

This choice comes with the advantage that $\sum_{k=0}^{M-1} W^2_{\text{rect},v,w}[k] = \sum_{k=0}^{M-1} W^2_{\text{sine},v,w}[k]$ guarantees that there is no over- or underrepresentation of energy in the overlap region of two windows as long as they share the same overlap length and type.

It has to be noted that window functions like Kaiser-Bessel-Derived (KBD) [20] [1] and certain sum-of-sines [21] amongst others share this property with the sine window and therefore can be used as well. This allows for the selection of the spectral window shape depending on their temporal response.

For the sine window we can analyze the temporal response in Figure 4. Obviously it decays much faster than the rectangular window and therefore produces less temporal leakage as is visible in the Figures 5/ 6 b and c respectively for the Hilbert and TNS envelope using the sine window. A TNS envelope like this allows for much better attenuation of the quantization noise where the actual signal is not active and thereby reduces pre-echo artifacts.

V. Evaluation

The evaluation of the new spectral windowing scheme is done using a simple experimental MDCT based codec system that allows for focussed evaluation of this aspect. A low-overlap window [22] of $M = 2048$ is used for temporal windowing to avoid potential TNS TDA artifacts [23]. The sampling frequency of the audio is set to 32000Hz. For the spectral windowing 4 different overlap lengths (0=rectangular window, 8, 16, 32) are used. The MDCT coefficients are grouped according to the AAC scale factor bands (SFBs)

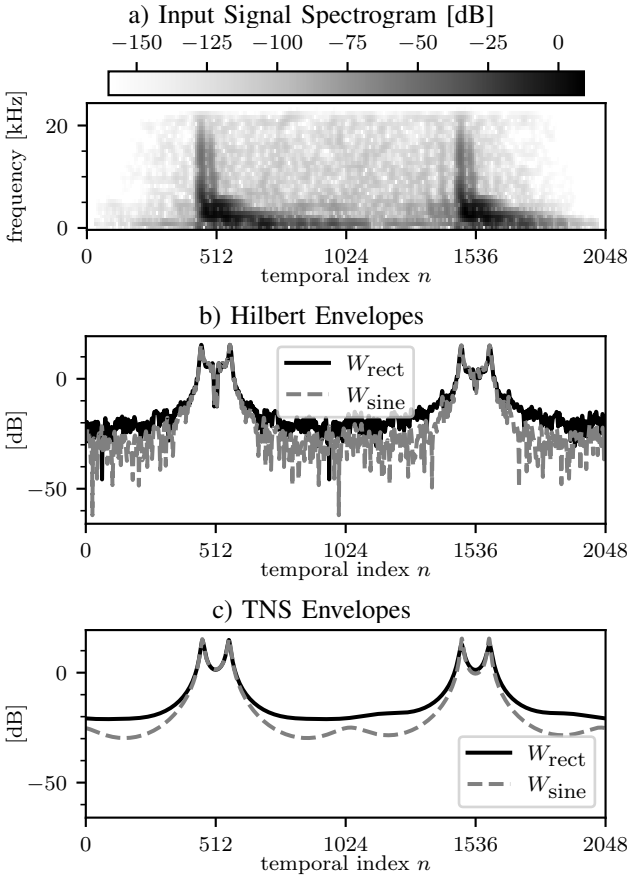

Fig. 6: Spectrogram of an electronic music ("Kalifornia" by "Fatboy Slim") input signal (a) to the MDCT, resulting Hilbert envelopes (b) and TNS envelopes (c). All parameters are the same as in Figure 5.

and quantized to an SNR (measured after TNS filtering at the encoder and before inverse TNS filtering at the decoder) of 12dB. Each frame may use up to 3 TNS filters of fixed frequency range. As in a lot of TNS implementations the TNS frequency range of each filter is divided into 3 sub-ranges for calculating 3 normalized sub-ACFs. These sub-ACFs are then averaged to obtain the final ACF for this range. This is done to avoid high energy bands dominating the temporal characteristics of the TNS filter. A TNS prediction gain threshold of 1.3 is used to switch on TNS. The TNS filters of order 8 are quantized as reflection coefficients in the arcsine domain using 4bit per coefficient.

In order to avoid that any of the systems have an advantage outside of the spectral windowing schemes, the following measures are applied. Entropy measurements on the quantized spectral coefficients are carried out to ensure all systems result in a comparable bitrate. As the prediction gain of the systems using the new spectral windowing scheme tends to be slightly higher, these systems are constrained to only activate TNS if the system using the spectral rectangular window would activate TNS as well.

The perceptual quality is evaluated using a MUSHRA [24] (Multi-Stimulus with Hidden Reference and Anchor) listening

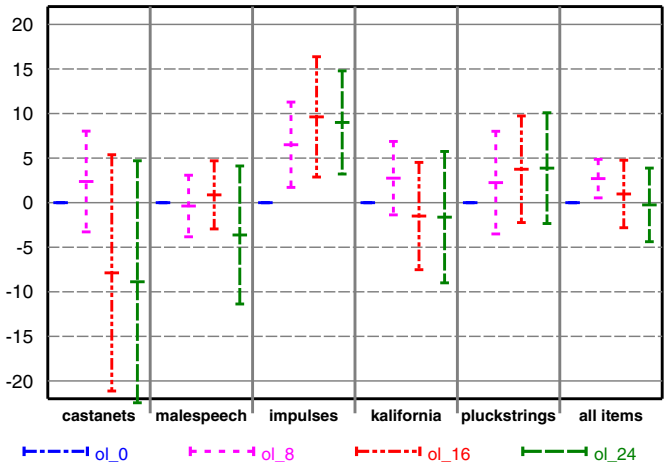

Fig. 7: Listening test results showing the MUSHRA difference scores of the 3 systems with spectral overlap ("ol") 8, 16 and 24 against the system with an overlap length of 0 (rectangular win.). Positive scores indicate an improvement compared to the rectangular window. 95% confidence intervals for the Gaussian distribution are shown. Reference and anchors were removed for clarity. 8 expert listeners participated using headphones.

test and the results are shown in Figure 7. From the results it is clear that among the new windowing scheme systems there is no clear winner. While a very large spectral overlap leads to improvements for very sharp transients with broad bandwidth (e.g. impulses), it sometimes leads to overly emphasized quantization noise due to sharper TNS envelopes (e.g. castanets) or worsened noise shaping due to reduced frequency resolution (e.g. male speech). Therefore the spectral overlap length of 8 is a good compromise and leads to a significant improvement of difference score over all items.

VI. Conclusion

A detailed mathematical description of TNS analysis and its impact on the noise shaping was given. Based on the theoretical results, an alternative spectral windowing scheme using overlapping, smooth windows for TNS analysis was proposed and its general benefit on coding quality was demonstrated. As the formulation in the GDFT domain allows for easy adoption to other DCT/DST types, application to the context of FDLP is straightforward. With the given theory further optimization of the spectral window parameters is possible.

Acknowledgment

The author would like to thank Bernd Edler and Jürgen Herre for their valuable feedback on the paper as well as the listening test participants for their time and effort.

References

[1] ISO/IEC, "Information technology – Coding of audio-visual objects – Part 3: Audio." International Standard 14496-3.

[2] ISO/IEC, "Information Technology—MPEG Audio Technologies — Part 3: Unified Speech and Audio Coding." International Standard 23003-3.

[3] ISO/IEC, "Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio." International Standard 23008-3.

[4] J. Princen, A. Johnson, and A. Bradley, "Subband/transform coding using filter bank designs based on time domain aliasing cancellation," in *ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 12, pp. 2161–2164, IEEE, 1987.

[5] A. Johnson and A. Bradley, "Adaptive transform coding incorporating time domain aliasing cancellation," *Speech communication*, vol. 6, no. 4, pp. 299–308, 1987.

[6] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–515, 2000.

[7] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, and M. Dietz, "ISO/IEC MPEG-2 advanced audio coding," *Journal of the Audio engineering society*, vol. 45, no. 10, pp. 789–814, 1997.

[8] F. Ghido, S. Disch, J. Herre, F. Reutelhuber, and A. Adami, "Coding of fine granular audio signals using High Resolution Envelope Processing (HREP)," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 701–705, IEEE, 2017.

[9] B. Edler, "Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen," *Frequenz*, vol. 43, no. 9, pp. 252–256, 1989.

[10] J. Herre and J. D. Johnston, "Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS)," in *Audio Engineering Society Convention 101*, Audio Engineering Society, 1996.

[11] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[12] J. O. Smith, *Spectral Audio Signal Processing*. `http://-ccrma.stanford.edu/~jos/sasp/`, accessed 2022. online book, 2011 edition.

[13] S. Ganapathy and H. Hermansky, "Temporal resolution analysis in frequency domain linear prediction," *The Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. EL436–EL442, 2012.

[14] Y. Wang, L. Yaroslavsky, and M. Vilermo, "On the relationship between MDCT, SDFT and DFT," in *WCC 2000-ICSP 2000. 2000 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000*, vol. 1, pp. 44–47, IEEE, 2000.

[15] Y. Wang, M. Vilermo, M. Väänänen, and L. Yaroslavsky, "Restructured audio encoder for improved computational efficiency," in *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[16] T.-W. Chang, C.-T. Chien, T. Chiou, Y.-H. Hsiao, H.-W. Hue, W.-C. Lee, C.-M. Liu, K.-Y. Peng, and C.-H. Yang, "Design of MPEG-4 AAC Encoder," in *Audio Engineering Society Convention 117*, Audio Engineering Society, 2004.

[17] H.-W. Hsu and C.-M. Liu, "Autoregressive modeling of temporal/spectral envelopes with finite-length discrete trigonometric transforms," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3692–3705, 2010.

[18] S. A. Martucci, "Symmetric convolution and the discrete sine and cosine transforms," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1038–1051, 1994.

[19] J.-L. Vernet, "Real signals fast Fourier transform: Storage capacity and step number reduction by means of an odd discrete Fourier transform," *Proceedings of the IEEE*, vol. 59, no. 10, pp. 1531–1532, 1971.

[20] L. D. Fielder, M. Bosi, G. Davidson, M. Davis, C. Todd, and S. Vernon, "AC-2 and AC-3: Low-complexity transform-based audio coding," in *Audio Engineering Society Conference: Collected Papers on Digital Audio Bit-Rate Reduction*, Audio Engineering Society, 1996.

[21] C. R. Helmrich, "On the Use of Sums of Sines in the Design of Signal Windows," in *Proceedings of DAFx-10, Graz*, 2010.

[22] E. Allamanche, R. Geiger, J. Herre, and T. Sporer, "MPEG-4 low delay audio coding based on the AAC codec," in *Audio Engineering Society Convention 106*, Audio Engineering Society, 1999.

[23] C.-M. Liu, H.-W. Hsu, and W.-C. Lee, "Compression artifacts in perceptual audio coding," *IEEE transactions on audio, speech, and language processing*, vol. 16, no. 4, pp. 681–695, 2008.

[24] R. International Telecommunication Union, "Recommendation ITU-R BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems (MUSHRA)," 2003.