Complex Wavelet Modulation Subbands for Speech Compression

Luneau, Jean-Marc; Lebrun, J.; Jensen, Søren Holdt

Published in:
Proceedings of Data Compression Conference, 2009

DOI (link to publication from Publisher):
10.1109/DCC.2009.52

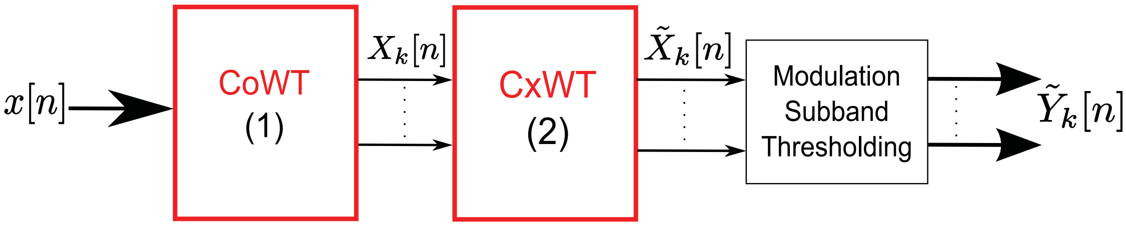
Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Complex Wavelet Modulation Subbands for Speech Compression

Jean-Marc Luneau[†], Jérôme Lebrun[‡], and Søren Holdt Jensen[†]
[†]*Department of Electronic Systems, Aalborg University, Denmark.*
[‡]*CNRS - I3S, UMR-6070, Sophia Antipolis, France.*
jml@es.aau.dk

Low-frequency modulation of sound carry essential information for speech and music [1]. They must be preserved for compression. The complex modulation spectrum has already been used for audio compression and is commonly obtained by spectral analysis of the sole temporal envelopes of the subbands out of a time/frequency analysis (Modified Discrete Cosine Transform combined with a Modified Discrete Sine Transform *e.g.* in [2]). However, amplitudes and tones of speech or music tend to vary slowly over time thus the temporal envelopes are often smooth and mostly of polynomial type. Processing in this domain usually creates undesirable distortions because only the magnitudes are taken into account and the phase data is often neglected. We remedy this problem with the use of a complex wavelet transform as a more appropriate envelope and phase processing tool. Complex wavelets carry both magnitude and phase explicitly with great sparsity and preserve well polynomials. Moreover an analytic Hilbert-like transform is possible with complex wavelets implemented as an orthogonal filter bank [3].



This figure illustrates the proposed method. x is the audio signal. X_k are the time/scale subbands (or scalogram) from the Continuous Morlet Wavelet Transform (CoWT) as defined by $CoWT_\sigma(x) = \int_{-\infty}^{+\infty} \Psi_\sigma(t)x(t)dt$ with $\Psi_\sigma(t) = C_\sigma \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2}(e^{i\sigma t} - e^{-\frac{1}{2}\sigma^2})$ and $\sigma = 10$. $\tilde{X}_k$ are the modulation scale/scale subbands from the Complex Wavelet Transform (CxWT) implemented via a non-redundant 3 band flexible orthogonal filterbank [3] with complex Daubechies wavelet filters of length 10. And finally $\tilde{Y}_k$ are the subbands after (hard or soft) thresholding and before quantization. By working in this alternative transform domain coined as “Modulation Subbands”, this transform shows very promising compression capabilities thanks to interesting sparsity properties and suggests new approaches for joint spectro-temporal analytic processing of slow frequency and phase varying audio signals.

References

[1] S. Greenberg, “On the origins of speech intelligibility in the real world,” *ESCA Workshop on robust speech recognition for unknown communication channels*, pp. 22–32, 1997.

[2] M. Vinton and L. Atlas, “A scalable and progressive audio codec,” *ICASSP*, vol. 5, 2001.

[3] J.-M. Luneau, J. Lebrun, and S. H. Jensen, “Complex wavelet modulation sub-bands and speech,” *ISCA ITRW on Speech Analysis and Processing for Knowledge Discovery, Aalborg*, vol. 1, 2008.

1068-0314/09 $25.00 © 2009 IEEE
DOI 10.1109/DCC.2009.52