Wavelets for speech signal feature extraction

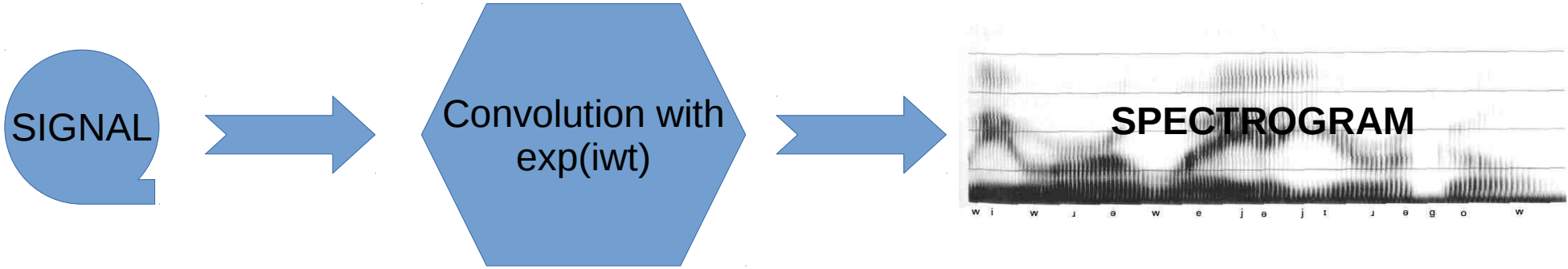
Wavelet Introduction

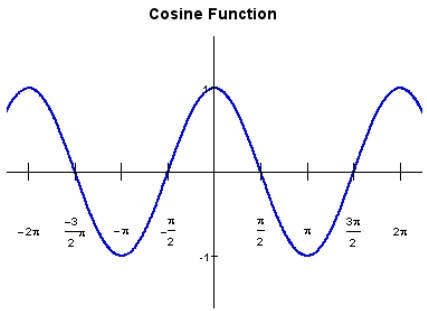
State of the art review

Open problems outline

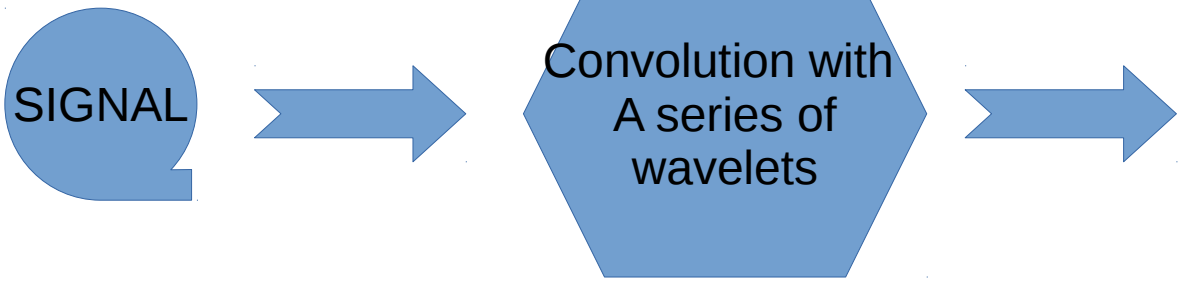
Sampled Continuous Wavelet Transform (SCWT)

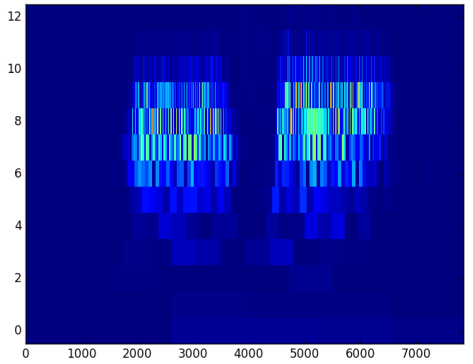
Wavelets' basics



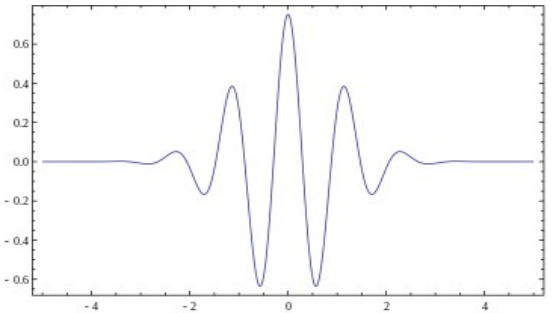


Wavelets' basics





SCALOGRAM



Wavelets' basics

- Good for:
  - Transient signal, because well localized in time
  - Noises, because of the scale aspect
  - Non-linear analysis
- Bad for:
  - Periodic signals, because of the localization in time.

➔ **It looks very good for unvoiced phonemes !!!**

State of the art

- Rather few research and papers:
  - Mainly about denoising [1][2][3][4][5]
  - Some papers about speech segmentation or detection [6][7][8][9]
  - Only a handful of papers phonemes/patterns oriented [10][11][12]

State of the art

- No analysis that satisfy me:
  - The main point of wavelets is for transient signals (unvoiced consonants)
  - Only two papers [10] from which a comparison between wavelets and Fourier approach on unvoiced sounds can be inferred (And for one it was not even the author intention!!!)
  - A hybrid system using wavelets for transients/noises and Fourier for resonants was not even suggested.

State of the art

- For example:
  - [10] Observes only isolated phonemes, but use dyadic WT instead of more refined WT.
  - [12] Uses a good WT but compares performances with MFCC on full words

Open problems outline

- Compare Continuous Wavelet Transform and “traditional” MFCC on an unvoiced sounds oriented recognition task
  - Like discriminating words with same vowels for example
  - Or isolating some plosives, like done in [10]
- Compare a hybrid system using Wavelet Transform for transients and “traditional” approach for resonants with both of the non-hybrid alternatives.
- Try to improve the Wavelet approach with a traditional preprocessing removing vowel influence

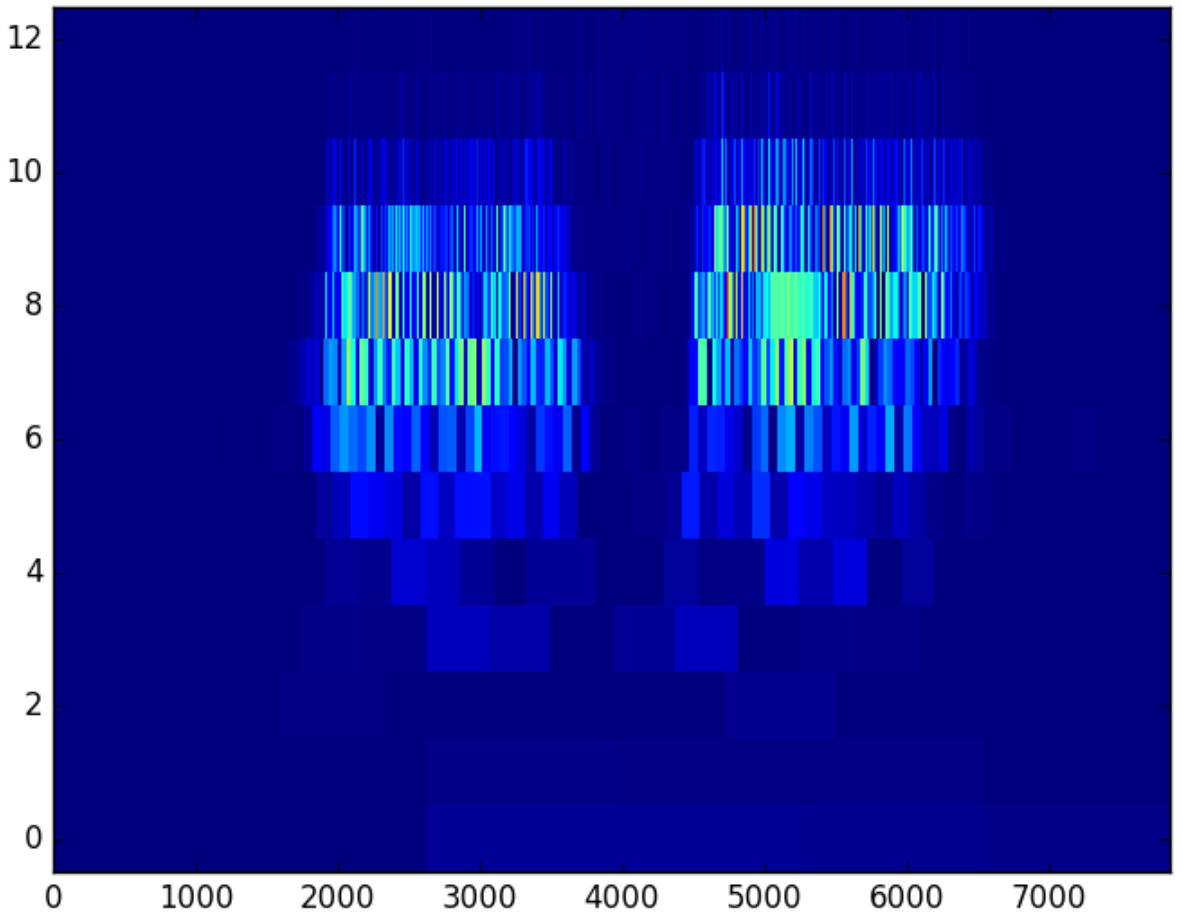
Sampled continuous wavelet transform (SCWT)

- Most of the papers are using the usual discrete, dyadic wavelet transform, which runs faster, and is available in most programing languages. (all but [12])
- However the scaling of the mother wavelet is dyadic (x2) → one coefficient (voice) per octave

  => Very very bad frequency resolution

  => Orthonormal

Sampled continuous wavelet transform (SCWT)



Sampled continuous wavelet transform (SCWT)

- An alternative is SCWT
- The scaling function is arbitrary (often 2^(m/V))
  - Not orthonormal anymore (can't reconstruct the signal, slower to compute). It is a highly redundant transformation.
  - Better frequency resolution (2 or 3 voices per octave seems wiser than one).

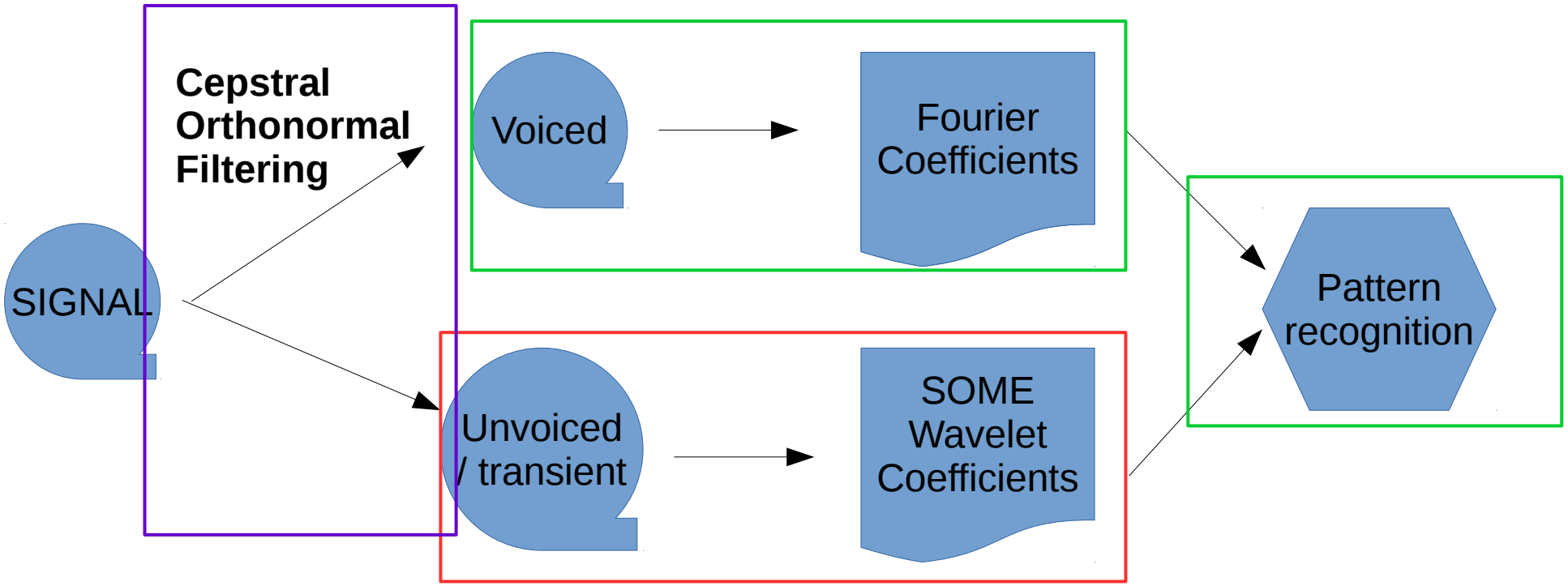
Sampled continuous wavelet transform (SCWT)

- Slow but:
  - Wavelets are used for recognizing transients which are rather short in time
  - In a hybrid system where MFCC takes out the load of the resonant part of the signal, the WT can be guided by the firsthand traditional approach to save time.

Sampled continuous wavelet transform (SCWT)

- Implementation are very scarce:
  - The python implementation only support mexican hat wavelet (bad for speech analysis).
  - I am currently trying R, with morlet wavelet (a complex wavelet).
  - Converting python numpy int16 to R dataframe is causing me some troubles.

Conclusion


Cepstral
Orthonormal
Filtering
SIGNAL
Voiced
Fourier
Coefficients
Unvoiced
/ transient
SOME
Wavelet
Coefficients
Pattern
recognition

References

[1] Hu, Yi, and Philipos C. Loizou. "Speech enhancement based on wavelet thresholding the multitaper spectrum." Speech and Audio Processing, IEEE Transactions on 12.1 (2004): 59-67.

[2] Bahoura, Mohammed, and Jean Rouat. "Wavelet speech enhancement based on the teager energy operator." Signal Processing Letters, IEEE 8.1 (2001): 10-12.

[3]Soon, Ing Yann, Soo Ngee Koh, and Chai Kiat Yeo. "Wavelet for speech denoising." TENCON'97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications., Proceedings of IEEE. Vol. 2. IEEE, 1997.

[4]LI, Yu, Feng-qin YU, and Shu-kai FAN. "Speech Enhancement Based on Improved Noise Variance Estimation in Wavelet Domain [J]." Audio Engineering 3 (2008): 022.

[5]Seok, Jong Won, and Keun Sung Bae. "Speech enhancement with reduction of noise components in the wavelet domain." Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on. Vol. 2. IEEE, 1997.

[6]Tan, Beng T., et al. "Applying wavelet analysis to speech segmentation and classification." SPIE's International Symposium on Optical Engineering and Photonics in Aerospace Sensing. International Society for Optics and Photonics, 1994.

[7]Alani, Ahmed, and Mohamed Deriche. "A novel approach to speech segmentation using the wavelet transform." Signal Processing and Its Applications, 1999. ISSPA'99. Proceedings of the Fifth International Symposium on. Vol. 1. IEEE, 1999.

[8]Ziolko, Bartosz, et al. "Wavelet method of speech segmentation." Signal Processing Conference, 2006 14th European. IEEE, 2006.

[9]Wendt, Christopher, and Athina P. Petropulu. "Pitch determination and speech segmentation using the discrete wavelet transform." Circuits and Systems, 1996. ISCAS'96., Connecting the World., 1996 IEEE International Symposium on. Vol. 2. IEEE, 1996.

[10] Farooq, O., and S. Datta. "Mel filter-like admissible wavelet packet structure for speech recognition." IEEE Signal Processing Letters 8.7 (2001): 196-198.

References

[11]Tufekci, Z., and John N. Gowdy. "Feature extraction using discrete wavelet transform for speech recognition." Southeastcon 2000. Proceedings of the IEEE. IEEE, 2000.

[12]Tan, Beng T., et al. "The use of wavelet transforms in phoneme recognition." Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on. Vol. 4. IEEE, 1996.

[13]Comparative Wavelet and MFCC Speech Recognition Experiments on the Slovenian and English SpeechDat2