Musical tones reduction using continuous wavelet transform for embedded speech recognition systems

J. Celerier

RITSUMEIKAN 音情報処理研究室 Acoustics & Signal Processing Labs

2013.7.20

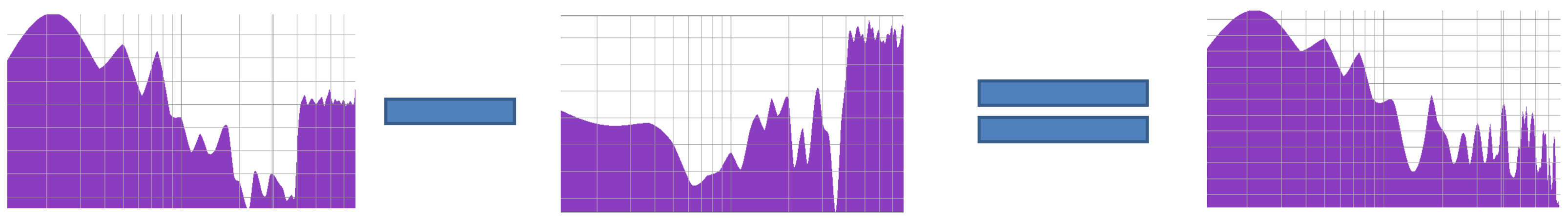
1. Introduction

Speech recognition

- Perturbations : **ambient noise**, distortion.
- Better if **real-time** processing : provides a smoother experience
- Many usages in **embedded devices** (smartphones, cars…)

Noise reduction

- Main method : **spectral subtraction** (SS) [1]
- Multiple optimizations possible: Iterative SS, Equal-Loudness-SS [2], Geometric approach [3]…



Problems with spectral subtraction

- Causes **musical tones** (MT) which disturbs speech recognition algorithms.
- MT are due to imperfect noise estimation, which causes over-subtraction.

2. Estimating musical tones

Time-frequency analysis

- Need to find a way to locate the musical tones.
- Useful : **continuous wavelet transform** (CWT) : allows recognition of sound characteristics. Gabor wavelet chosen.

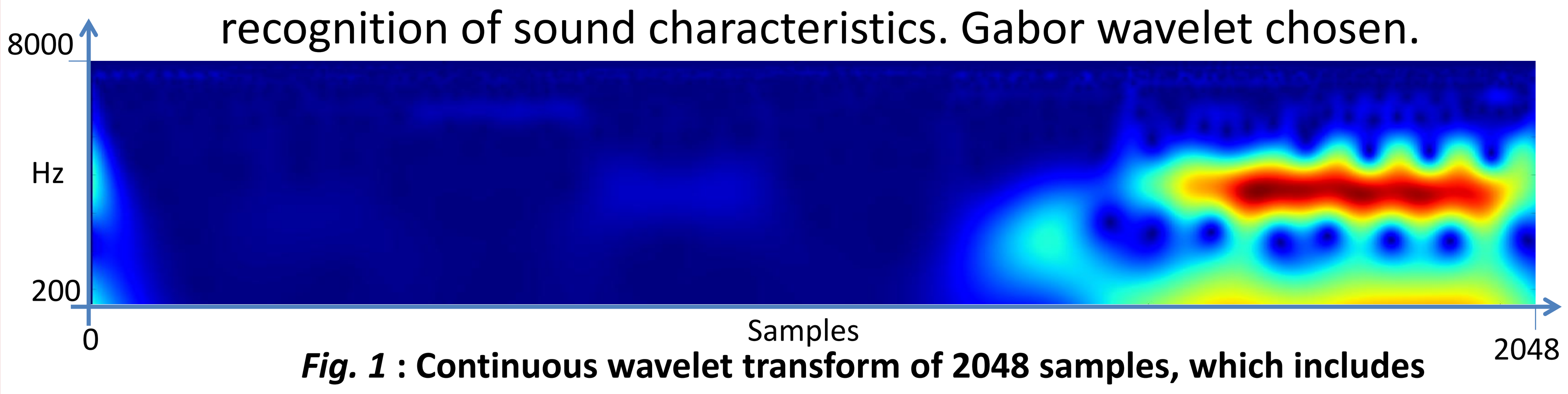


Fig. 1 **: Continuous wavelet transform of 2048 samples, which includes both vocal features (right) and musical tones (left & center).**

Heuristics & recognition

- Needed to **differentiate** between MT and vocal features.
- Heuristics : length of features, constancy of the MT pitch (stays the same during one FFT duration), research of short bursts.
- Shape recognition in CWT : uses **ceiling** and **contour detection**.

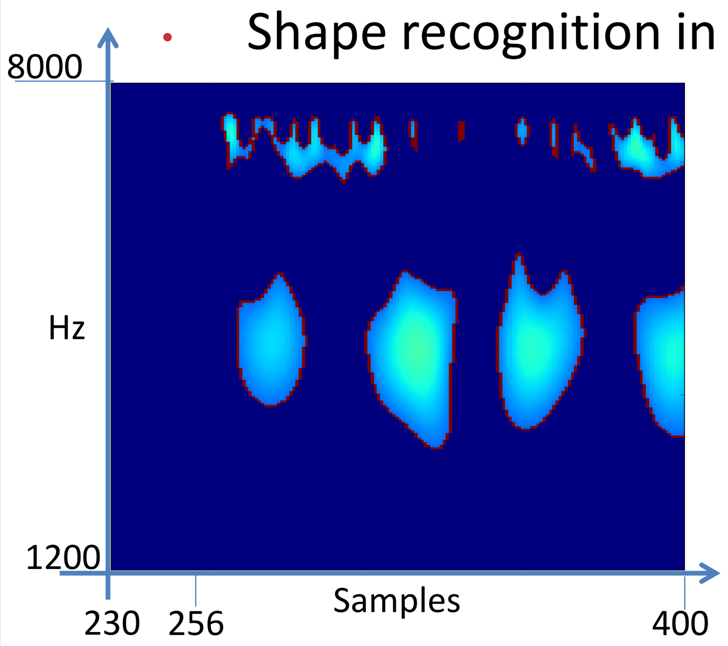


Fig. 2 **: Detected musical tones areas (upper left corner of previous plot).**

Once areas are computed, computation of the over-subtracted power spectrum takes place:

- Removal of areas with a large frequency span.
- Computation of the average frequency and power of an area.
- Computation of the corresponding FFT bin for this FFT frame.

3. Removing musical tones

Directly removing from CWT plane

- Spectral-temporal subtraction.
- Might cause artifacts because of surroundings of detected areas.
- Need to perform inverse-CWT : **too expensive** in processing power.

Per-frame noise reestimation

- Decrease frequency bins of noise estimation where musical tones are present, and **re-perform a spectral subtraction** on original noisy signal with modified noise estimation.

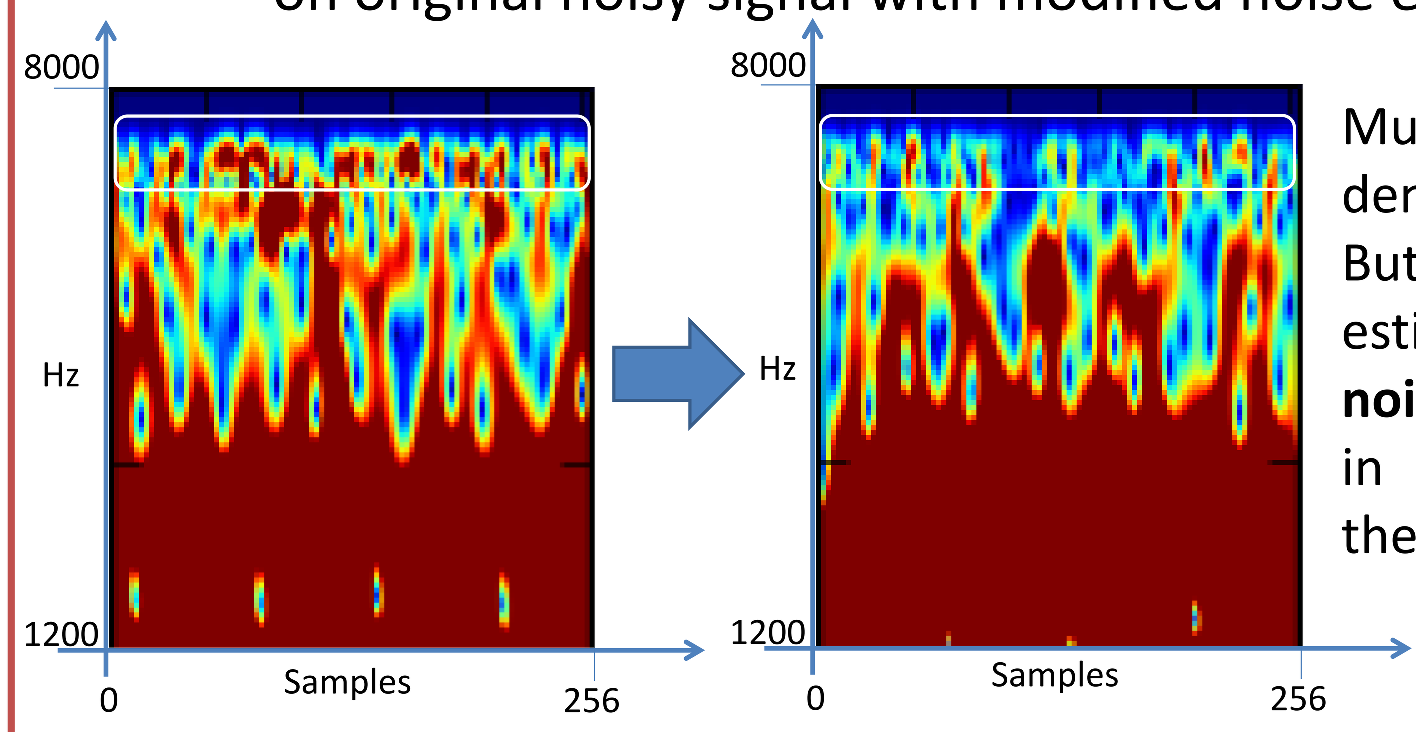


Fig. 3 **: No reestimation** ***Fig. 4*** **: Reestimation**

Musical tones density decreases. But imperfect estimation causes **noise resurgence** in other bins of the spectrum.

4. Evaluation

Common evaluation parameters

- Signal evaluation [2]
  - Noise reduction ratio (NRR)
  - Speech distortion ratio (SDR)

$$\gamma = \frac{\sum_{k=0}^{N-1} |x_k|}{\sum_{k=0}^{N-1} |y_k|} \qquad \mathrm{NRR} = 10 * \log_{10}\left(\frac{\sum_{k=0}^{N-1} z_k^2}{\sum_{k=0}^{N-1} y_k^2}\right)$$

x : clean signal
y : processed signal
z : noisy signal

$$\mathrm{SDR} = 10 * \log_{10}\left(\frac{\sum_{k=0}^{N-1} x_k^2}{\sum_{k=0}^{N-1}(x_k - \gamma * y_k)^2}\right)$$

Experimental parameters	
Noise database	JEIDA
Vocal database	TIMIT
Sampling freq.	16 000 Hz
Bit depth	16 bit
CPU	2.5 GHz

Comparison of NRR and SDR with standard method

Fig. 5 : Reaction to signal-to-noise ratio

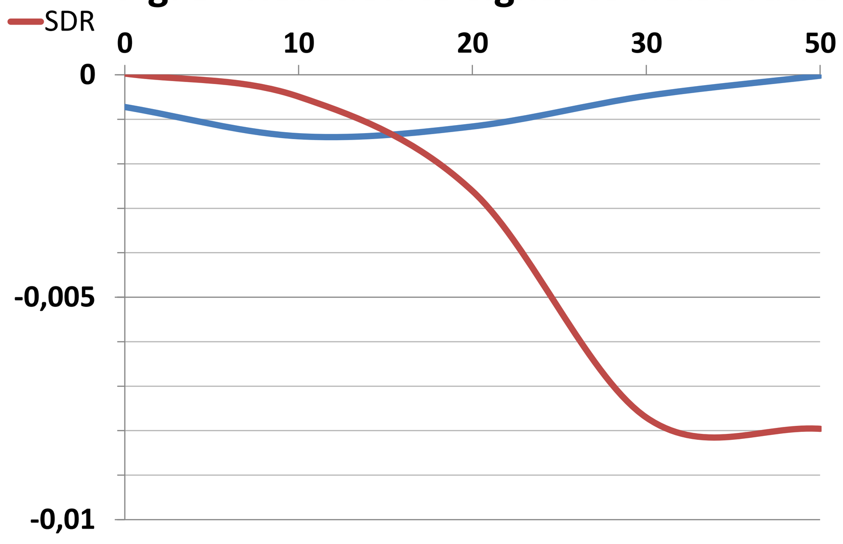
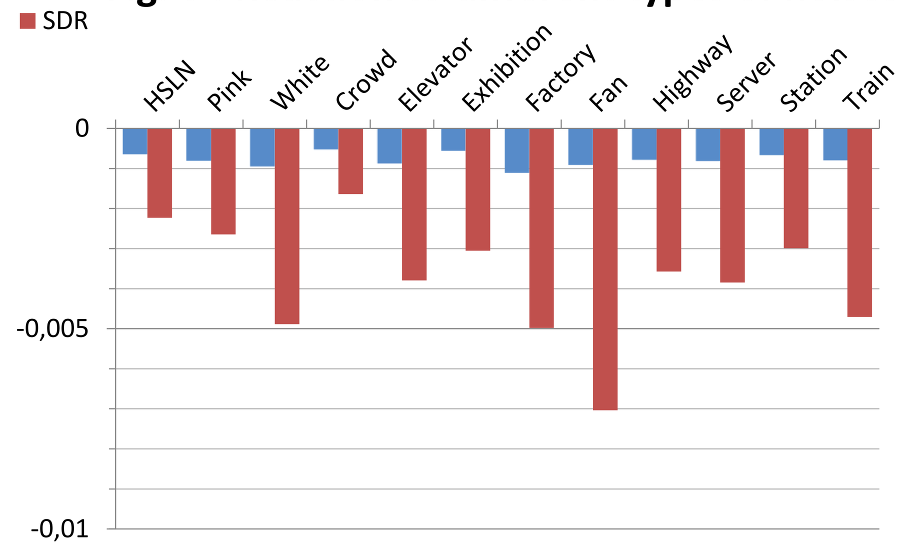


Fig. 6 : Reaction to different types of noises



Proposed method slightly decreases both NRR and SDR, because noise is added again due to poor estimation.

Time measurements (Processing of 10 seconds of speech)

- 1 iteration without CWT : 0.1 seconds
- 1 iteration with CWT : 20 seconds
  - Unsuitable for real-time.

5. Conclusion

- Not on par with other musical tones reduction methods ([4], [5]) : too slow and nearly irrelevant results.
- Room for improvement :
  - Using fast FFT library (ex. : FFTW) to increase speed of CWT.
  - Make heuristics more precise.
- Future work :
  - Applying wavelet estimation only in high frequencies to improve efficiency.
  - Try to find minimum wavelet transform precision that allows musical tones detection, in order to reduce processing time.

References

[1] Boll, S. : Suppression of acoustic noise in speech using spectral subtraction.
[2] Horii, K., Fukumori, T., Nakayama, M., Nishiura, T., Yamashita, Y. : Musical tone reduction for sound-quality improvement by weighted iterative spectral subtraction in real noisy environments.
[3] Lu, Y. : A geometric approach to spectral subtraction.
[4] Inoue, T. : Theoretical Analysis of Musical Noise in Generalized Spectral Subtraction Based on Higher Order Statistics.
[5] Goh, Z. : Postprocessing Method for Suppressing Musical Noise Generated by Spectral Subtraction.