Enhancement of reverberant Speech using Kurtosis of LP Residual

Vikrant Tomar

McGill ID: 260394445

telecommunications & signal processing laboratory

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

April 10, 2010

Outline

Next

Scope and Motivation

- Ideally, speech communication systems use direct signal of a source
- In real life scenarios, major challenges originates from the degradation in capturing speech in a confined space:
  - Background noise
  - **Reverberation**
  - Other interferences
- Linear Prediction residual is a very good metric of reverberation in speech.

Reverberation

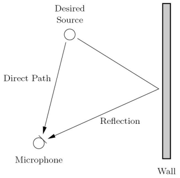
- Reverberation: Presence of delayed and attenuated copies of source signal in the received signal.
- Cause: Reflections from various surfaces in the room.
- Received Signal is the clean speech convolved with the Room Impulse Response (RIR), and added ambient noise:

$$x(n) = s(n) * g(n) + w(n)$$

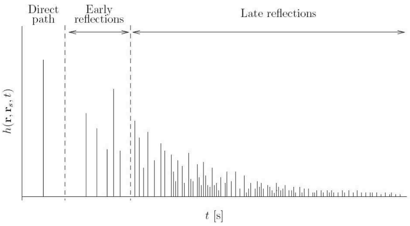
- Measure: **Reverberation time ($\mathbf{T_{60}}$)**: Time required for a sound to drop 60 decibels or to decay to a value one millionth of its original intensity.

Acoustic Environment of a Room

Example (The simplest case of single reflective surface)



A typical Room Impulse Response.



Next

Dereverberation

Aim To get clean speech out of reverberated signal using deconvolution *(in some sense)*.

Dereverberation

Aim To get clean speech out of reverberated signal using deconvolution *(in some sense)*.

Classification

- General Signal $\longleftrightarrow$ Speech Signal
- Based on Enhancement:
  Human Perception $\longleftrightarrow$ Automatic Speech Recognition (ASR)
- Single Channel $\longleftrightarrow$ Multi-Channel

Approaches

General Signal

No *a priori* knowledge about the signal.

Speech Signal

Various properties of speech and language can be utilized, such as: voiced/unvoiced speech, harmonic structure, Model of speech production, *etc.*

Approaches

For Better Human Perception

- Improve intelligibility
- Try to get a nicer, nearer sound.
- Rather leave some reverberation than causing spectral distortion
- Application: Hands-free Telecommunication

Approaches

For Better Human Perception

- Improve intelligibility
- Try to get a nicer, nearer sound.
- Rather leave some reverberation than causing spectral distortion
- Application: Hands-free Telecommunication

For Better ASR:

- Get rid to as much reflections as possible.
- A nice sound, in terms of human perception, is not aspired.

Linear Prediction and residual

- Source-Filter Speech Production AR Model:
  - Excitation source (Glottal Pulse train, noise)
  - Vocal Tract (Modelled as an All-pole filter by linear prediction)
- Speech Signal, $s(n) = u(n) - \sum_{k=1}^{l} h(k)s(n-k)$
- Predicted Signal, $\tilde{s}(n) = \sum_{k=1}^{l} a(k)s(n-k)$
- LP residual, $e(n) = s(n) - \tilde{s}(n)$
- If speech signal were to be truly the response of an all-pole model $\rightarrow$ exact predictability at all instants except the excitation instant, *i.e.*,

  $$e(n) = u(n);\ \text{For } a(n) = h(n)$$
- Hence, LP residual essentially represents the excitation signal for voiced speech.

Linear Prediction and Reverberation

- Reverberation mainly affects the excitation signal $\Rightarrow$ LP residual
- Many different approaches to manipulate the LP residual.
    - Apply an adaptive weight function to the LP-residual, [5]. (emphasizes regions with high signal-to-reverberation ratio.)
    - Apply a Code Excited Linear Prediction (CELP) post-filter to the LP-residual.
    - **Exploit the higher order statistical properties of LP-residual, e.g., Kurtosis [2, 4]**.

Next

Maximum Kurtosis based Adaptive Filtering: An Overview

- Control the LMS-like adaptive filter by the "Kurtosis" of the LP residual
- Make use of *a priori* knowledge that the signal to be used is Speech.
- Speech Enhancement for human perception (useful for applications related to hearing aids).
- Single/Multi-channel implementation
- Exploits the properties of cumulants, and speech: Higher order statistics of a Gaussian distribution are zero, hence removes the Gaussian Noise.

What is Kurtosis?

- Etymological meaning: Bulging
- A measure of the "peakedness" of the probability distribution of a real-valued random variable.

$$\Gamma = \frac{\kappa_4}{\kappa_2^2} = \frac{\mu_4}{\sigma^4} - 3 \tag{1}$$

where:

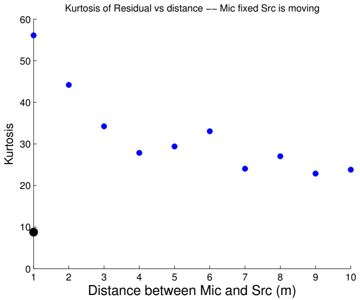
  - κ_n := n-th cumulant
  - μ_n := n-th central moment, $:= E[(X-\mu)^n]$
  - σ := second central moment (= std. div.)

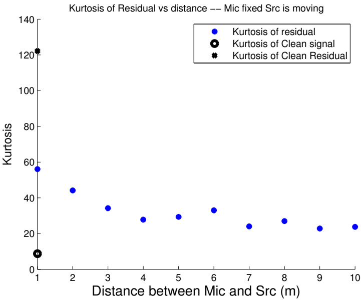
- Kurtosis of the normal distribution is 3; 3 is subtracted for normalization.
- Note that cumulants show linear behaviour.

a priori knowledge!

- For clean voiced speech, LP residuals have strong peaks corresponding to glottal pulses.
- For reverberated speech, such peaks are spread in time.
- Implying that, Kurtosis of LP residuals reduces as reverberation increases.
- It has been empirically proved that the Kurtosis of LP residual is a reasonable reverberation metric, [2].

Kurtosis of LP residual: Some results

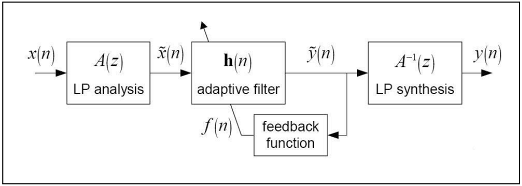




Simulated Room Environment with following details:
$4 \times 13 \times 4\ m^3$
Mic: Fixed at [2, 2, 2]
Source: Moving from [2, 3, 2] to [2, 12, 2] (10 points)
Reverberation time: $T_{60} = 0.4$ seconds
Note: Identical results if source is fixed and mic is moving.

Implementation - 1

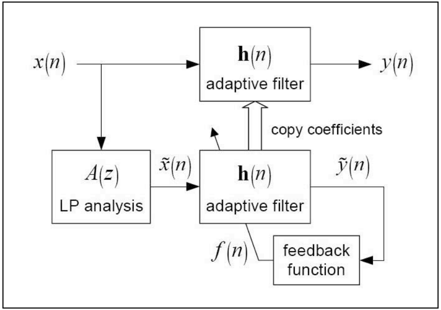
- Develop online adaptive gradient ascent algorithm, that maximizes the LP residual kurtosis.



- Issue: Only valid under the assumption that the LP-coefficients are unaffected by the multipath effects of the room (Holds only in spatially averaged sense).

Implementation - Final

- Reliance on the assumption is removed by using an additional filter with identical coefficients.



Feedback Function

Cost function = Kurtosis of Residual

$$J(n) = \frac{E[\tilde{y}^4(n)]}{E^2[\tilde{y}^2(n)]} - 3 \tag{2}$$

Differential of Cost function

$$\frac{\delta J}{\delta \vec{\mathbf{h}}} = \left(\frac{\left(E[\tilde{y}^2]y^2 - E[\tilde{y}^4]\right)\tilde{y}}{E^3[\tilde{y}^2]} \right) \tilde{\mathbf{x}} = f(n)\tilde{\mathbf{x}}(n) \tag{3}$$

Feedback function

$$f(n) = \left(\frac{\left(E[\tilde{y}^2]y^2 - E[\tilde{y}^4]\right)\tilde{y}}{E^3[\tilde{y}^2]} \right) \tag{4}$$

Update Equation

- For single channel implementation:

$$\vec{\mathbf{h}}(n+1) = \vec{\mathbf{h}}(n) + \mu f(n)\tilde{\mathbf{x}}(n) \tag{5}$$

- For multi channel implementation:

$$\vec{\mathbf{h}}_c(n+1) = \vec{\mathbf{h}}_c(n) + \mu f(n)\tilde{\mathbf{x}}_c(n) \tag{6}$$

  Each channel is independently adapted using the same feedback function.

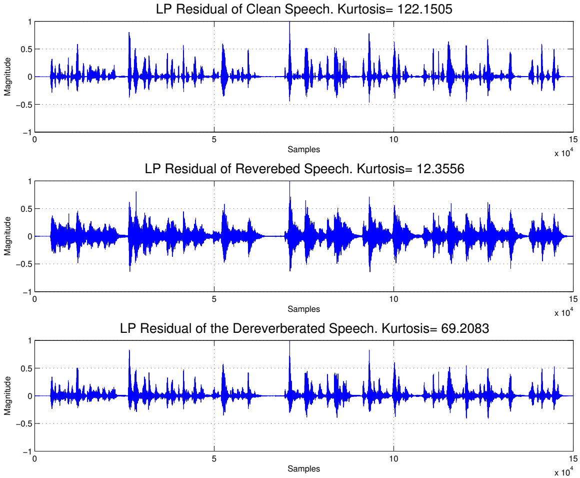
- $E[\tilde{y}^2(n)]$ and $E[\tilde{y}^4(n)]$ are estimated recursively:

$$E[\tilde{y}^2(n)] = \beta E[\tilde{y}^2(n-1)] + (1-\beta)\tilde{y}^2(n) \tag{7}$$

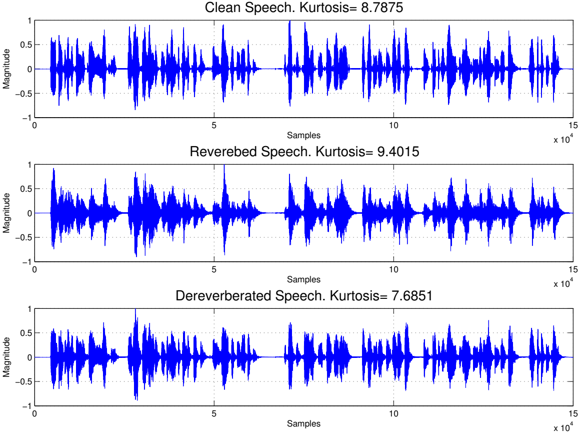
$$E[\tilde{y}^4(n)] = \beta E[\tilde{y}^4(n-1)] + (1-\beta)\tilde{y}^4(n) \tag{8}$$

Experiments and Results

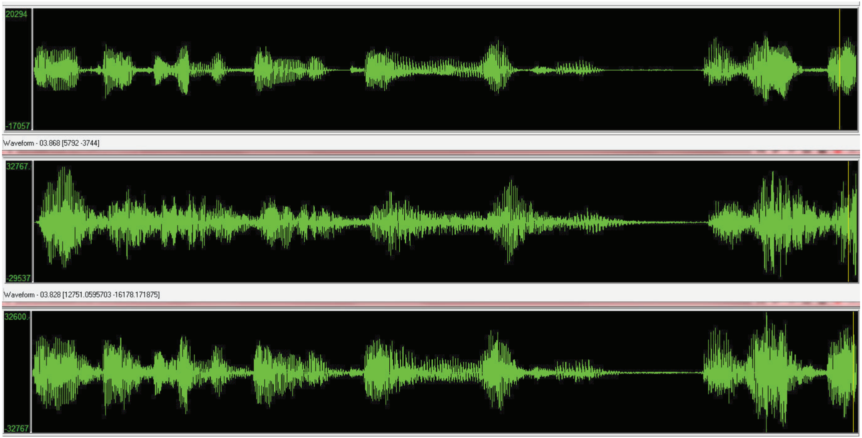
Case 1: Clean Speech $F_s = 8000$ Hz; Room: Dim: $4 \times 4 \times 4$ m^3; Mic: [2,3,2], Src: [2,2,2]; T_{60}=0.7 Sec



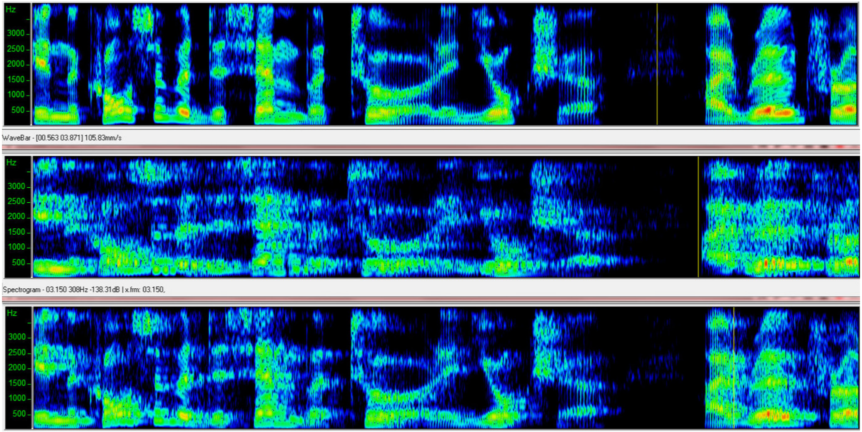
Waveform – Matlab



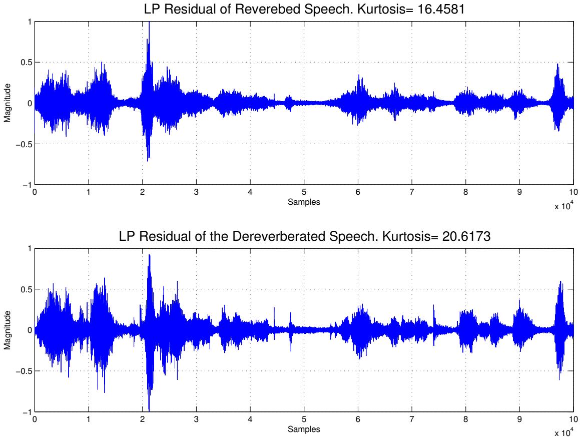
Waveform – Wavesurfer



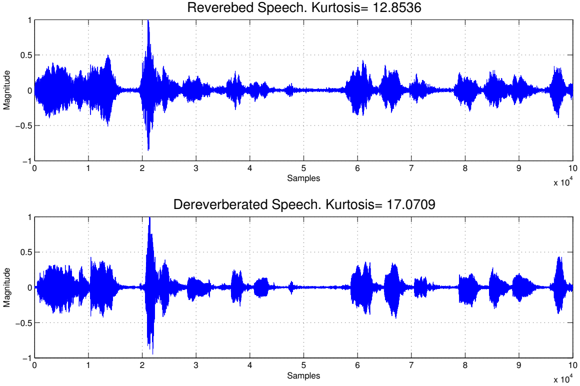
Spectrogram



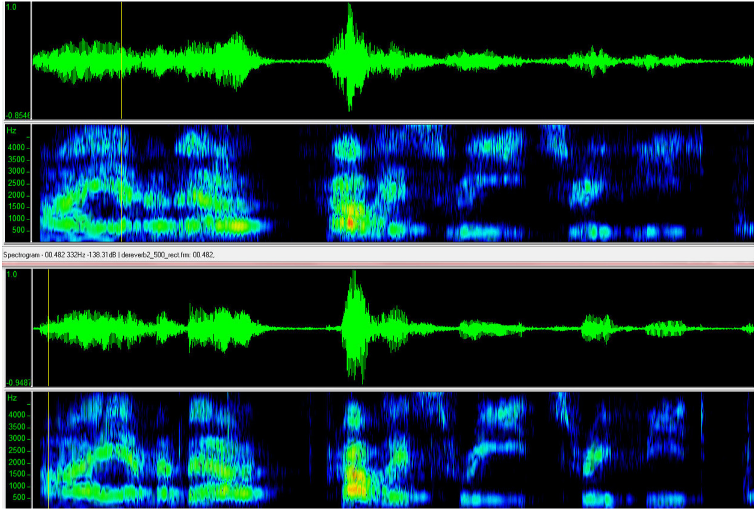
Case 2: Blind Dereverberation. No info about Room or Clean Speech. $F_s = 16000$ Hz;



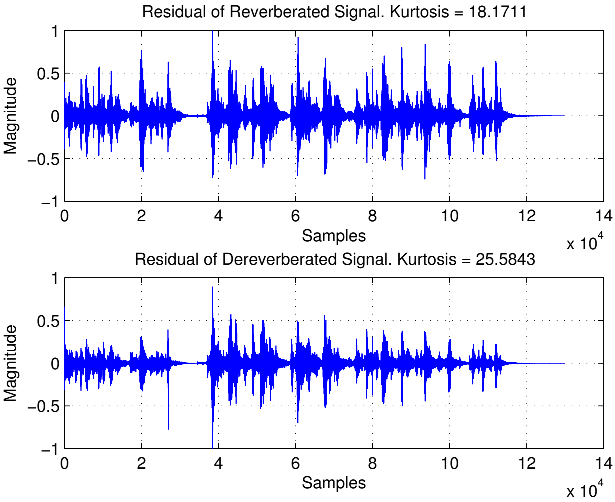
Waveform – Matlab



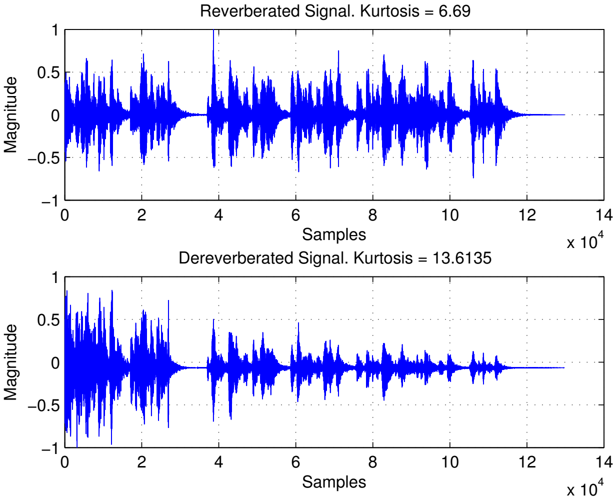
Spectrogram



Case 3: Blind online Dereverberation. No info about Room. $F_s = 8000$ Hz, Room Impulse response taken from AIR databese, Aachen University, Germany [3]



Waveform – Matlab



Next

Conclusions and Future Work

- Kurtosis of LP residual is an effective measure of reverberation in speech.
- Significant improvement in reverberated speech was noticed.
- However, does NOT work very well with unvoiced discontinuities in speech.

Future Work

- Authors in [1] claims that an average over several spatially distributed microphones can provide potentially better results.
- Time domain implementation is prone to slow or no convergence because of all variance in the eigenvectors of autocorrelation matrices of the input signal.
- A subband adaptive method is promoted in [2] which is more robust to noise.

References

Nikolay D. Gaubitch, Patric Naylor, and Darren Ward.
On the use of Linear Prediction for Dereverberation of Speech.
In *Intl. Workshop on Acoustic Echo and Noise Cancellation*, Kyoto, Japan, 2003.

B. W. Gillespie, D. A. F. Florencio, and H. S. Malvar.
Speech dereverberation via maximum-kurtosis subband adaptive filtering.
In *ICASSP '01: IEEE International Conference on Acoustics, Speech, and Signal Processing 2001*, volume 6, pages 3701–3704. IEEE.

Marco Jeub, Magnus Schaxfer, and Peter Vary.
A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms.
In *16th International Conference on Digital Signal Processing (DSP)*, Santorini, Greece, 2009.

O Tanrikulu and A G Constantinides.
Least-mean Kurtosis: a novel higher-order statistics based adaptive filtering algorithm.
Electronic Letters, 30(3):189–190, Feb 3, 1994.

B. Yegnanarayana and P. Satyanarayana Murthy.
Enhancement of of reverberant speech using LP residual Signal.
IEEE Transactions on Speech and Audio Processing, 8(3):267–281, May 2000.

Next

Thank You!

Cumulants

- The cumulants k_n of a random variable X are defined by the cumulant generating function, which is logarithm of the moment-generating function.
- $g(t) = log(E[e^{tX}]) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!} = \mu t + \sigma^2 \frac{t^2}{2} + ...$
- Cumulants are then given by derivatives at zero of g(t).
  - $k_1 = \mu = g'(0)$
  - $k_2 = \sigma = g''(0)$, and so on...