Blind dereverberation of single channel speech signal based on harmonic structure

Tomohiro Nakatani and Masato Miyoshi

Speech Open Lab.,

NTT Communication Science Labs.,

NTT Corporation

Background & Purpose

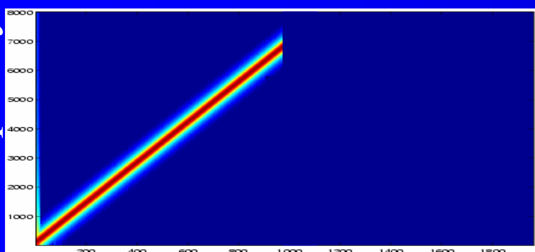
- Reverberation degrades speech recognition
  - Dereverberation is required as pre-processing
- Existing method for dereverberation
  - Multi-mic.: beam-forming, inverse filtering
  - Single-mic.: Cepstrum based, LPC residue based
    → insufficient for long reverberation
- Proposed method
  - Single channel blind dereveberation
  - Estimate inverse filter based on harmonics

Basic idea

Source signal

$X(\omega)$

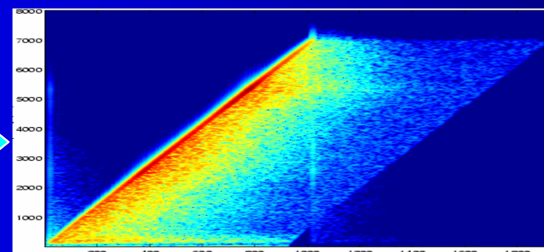
frequency



time

Reverberant signal

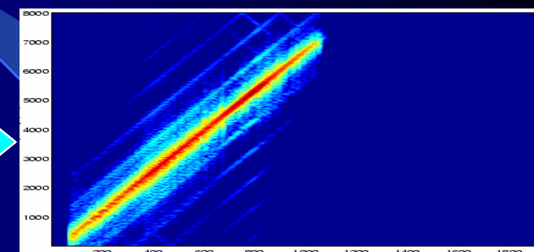
$Y(\omega) = H(\omega)X(\omega)$



Transfer function : H

Approximated direct signal

$X'(\omega)$



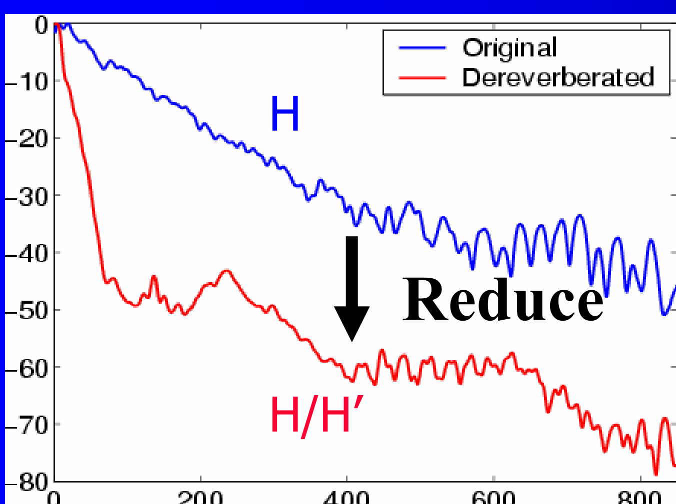
Extract a dominant sinusoid

Inverse transfer function

$1/H'(\omega) = X'(\omega)/Y(\omega)$

Utilize as inverse filter

(dB)



Reverberation curve (time)

Speech modeling

Source signal

$X = X_h + X_n$ → H → Reverberant signal $Y = HX$ → Harmonics filter → Approximated direct signal $\hat{X}'_h$

$E(\hat{X}'_h / Y)$

Average inverse transfer function

X_h : voiced (harmonics)

X_n : unvoiced (non-harmonics)

Models

Reverberant signal

$$Y = HX$$
$$= (D + R)X$$
$$= DX_h + (RX_h + HX_n)$$

Approximated direct sound

$$\hat{X}'_h = DX_h + (\hat{R}X_h + \hat{N})$$

D : Direct signal component

R : reverberant signal component

Principle of dereverberation

- Average inverse transfer function $E(\hat{X}'_h / Y)$

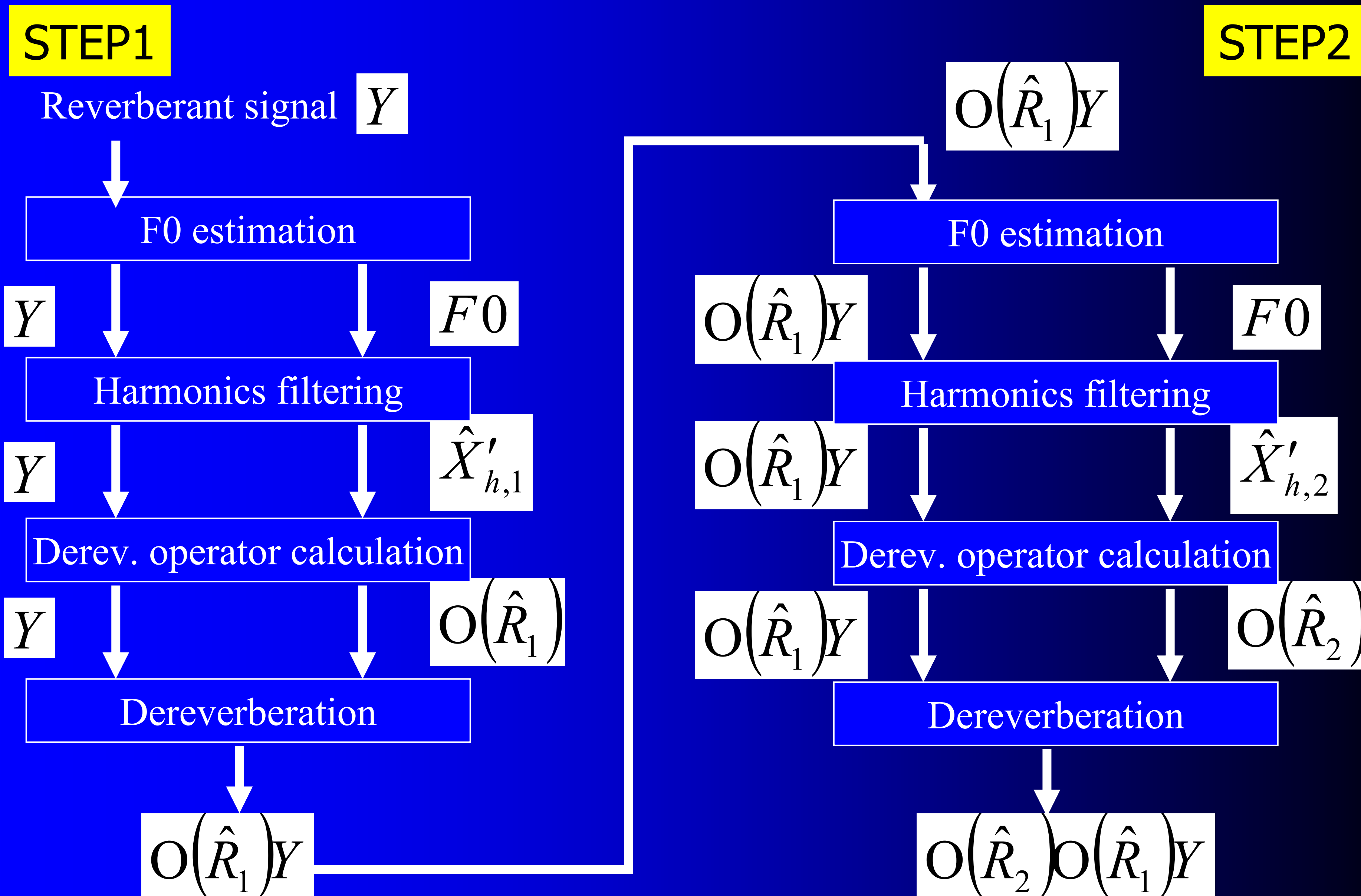
$$E\left(\hat{X}'_h / Y\right) = \frac{\left(D + \hat{R}\right)}{H} E\left(\frac{1}{1 + X_n / X_h}\right) + E\left(\frac{1}{1 + \left(Y - \hat{N}\right) / \hat{N}}\right)$$

$$\approx \boxed{\frac{\left(D + \hat{R}\right)}{H}} P\left(\left|X_h\right|^2 > \left|X_n\right|^2\right)$$

- Dereverberation operator $\mathrm{O}\left(\hat{R}\right) = \left(D + \hat{R}\right) / H$
  - Reduce reverberation by multiplying Y by $\mathrm{O}\left(\hat{R}\right)$

$$\mathrm{O}\left(\hat{R}\right) Y = \left(D + \hat{R}\right) X$$

Processing flow



Harmonics filter

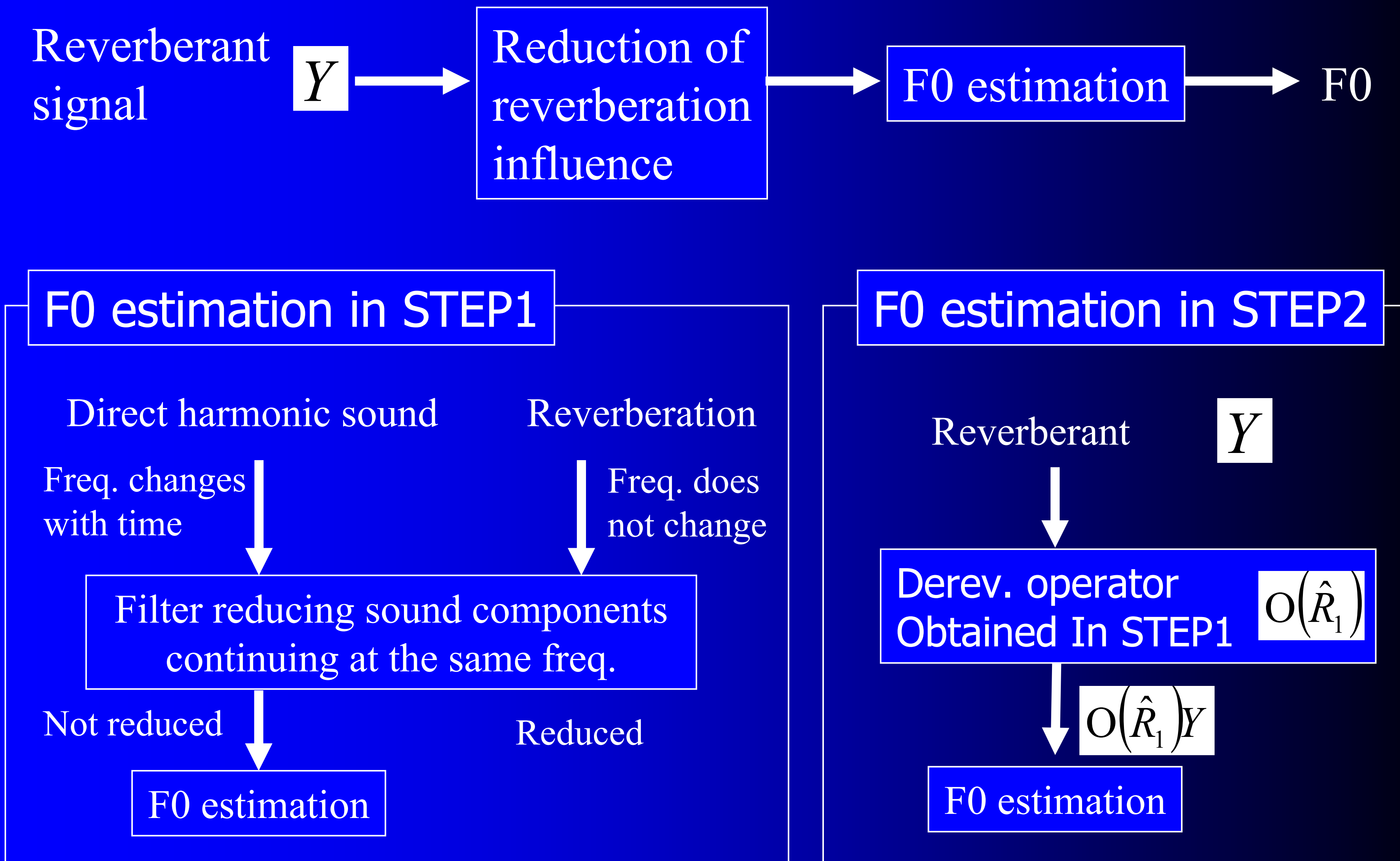
1. Estimate fundamental freq. $f_0(n)$ at each freme n
2. Extract phase $\hat{\phi}_k(n)$ and magnitude $\hat{a}_k(n)$ of the k-th harmonic from a DFT bin corresponding to $kf_0(n)$
3. Synthesize harmonic sound by adding sinusoids

$$\hat{x}(t) = \sum_n \sum_k w(t - \tau_n)\hat{a}_k(n)\sin\left(2\pi k f_0(n)t + \hat{\phi}_k(n)\right)$$

⇒ Enhance direct harmonic sound reducing other components

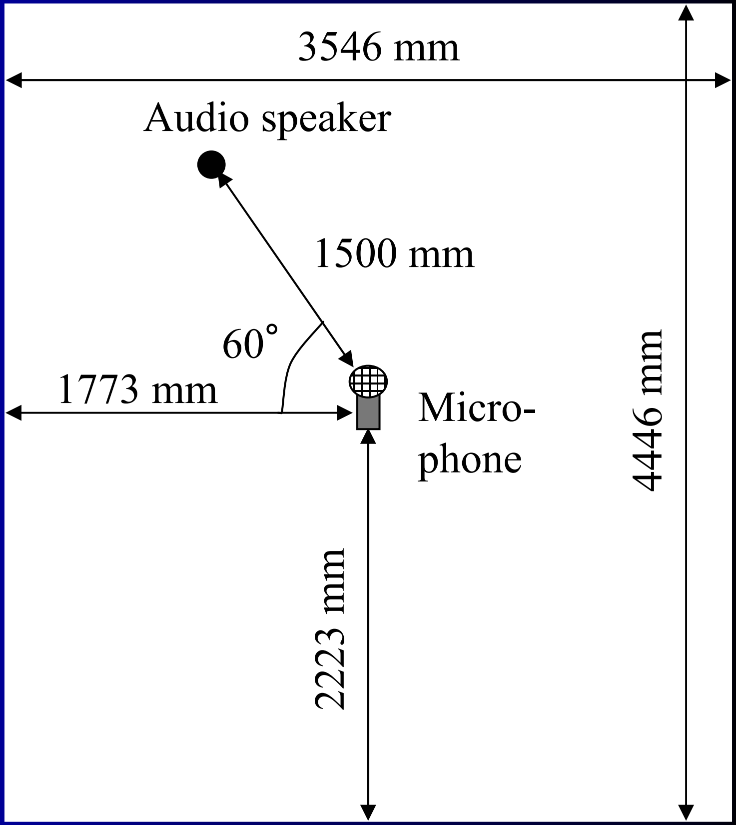
But it requires robust fundamental frequency (F0) estimation in the presence of reverberation

Robust F0 estimation with reverberation



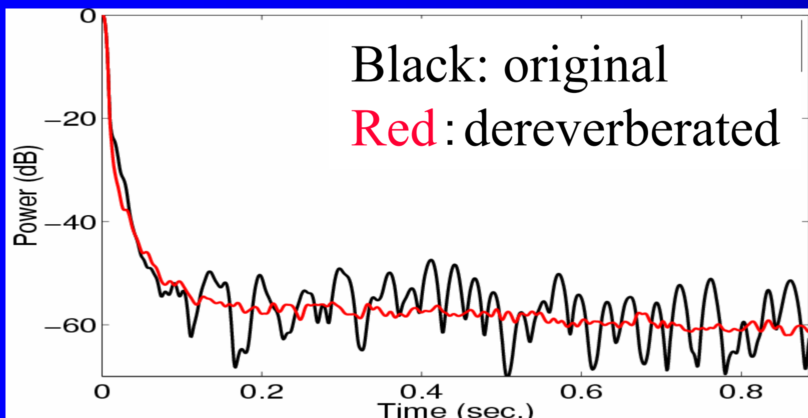
Experiment condition

- Source signal
  - ATR word DB (12kHz,16bit)
  - Female: FKM（5240 words）
  - Male: MAU（5240 words）
- Impulse response
  - Measured with the reverberation time of 0.1, 0.2, 0.5, and 1.0 sec.
- Reverberant signal
  - Synthesized by convolving source signal with impulse responses

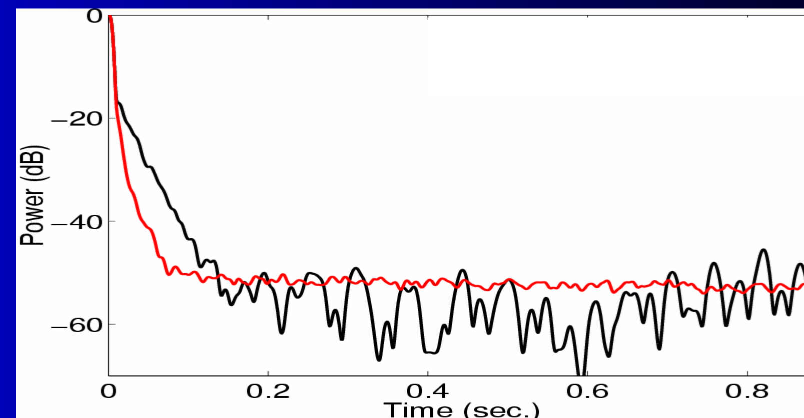


Impulse response measurement

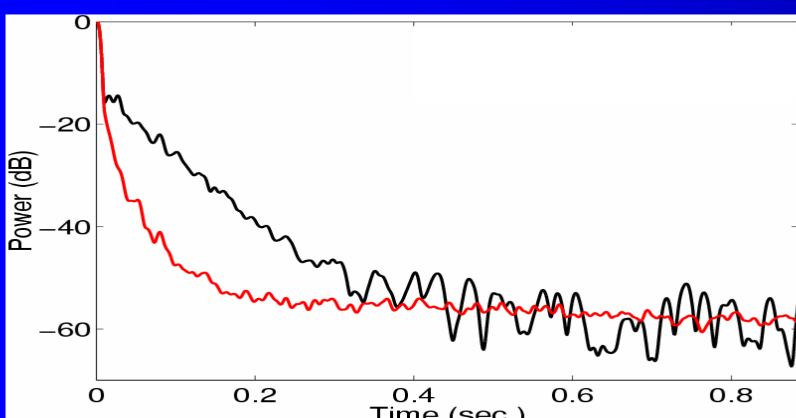
Reverberation curves (female)



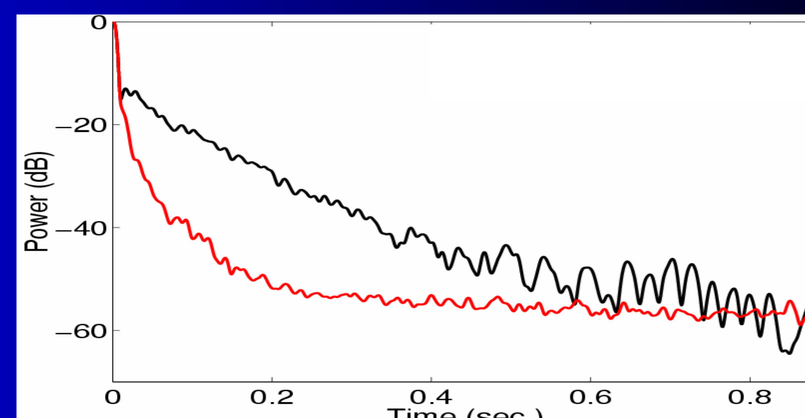
Reverberation time: 0.1 sec



Reverberation time: 0.2 sec

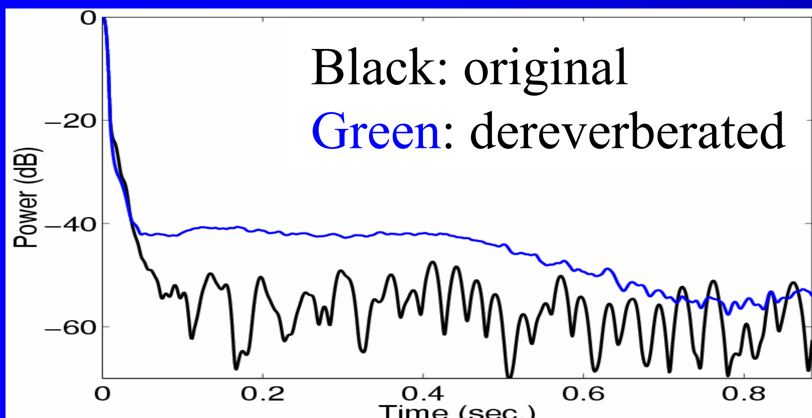


Reverberation time: 0.5 sec

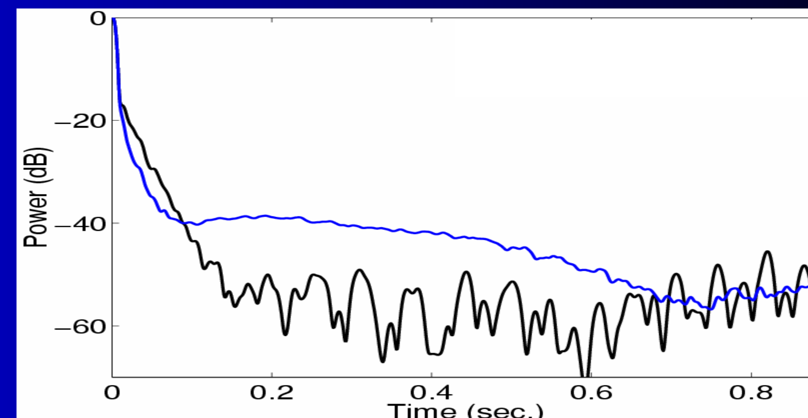


Reverberation time: 1.0 sec

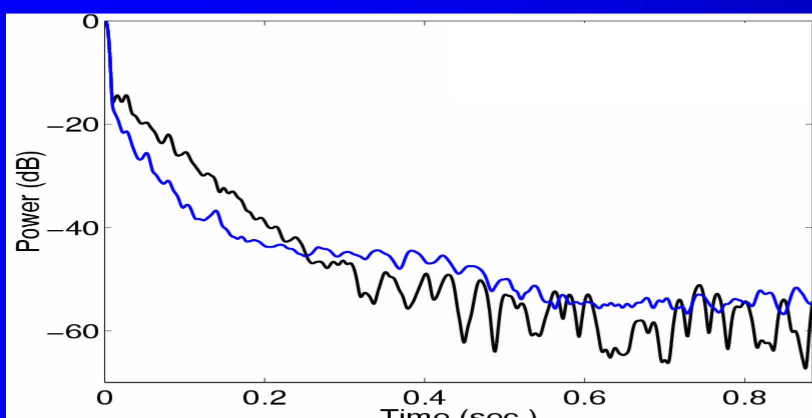
Reverberation curves (male)



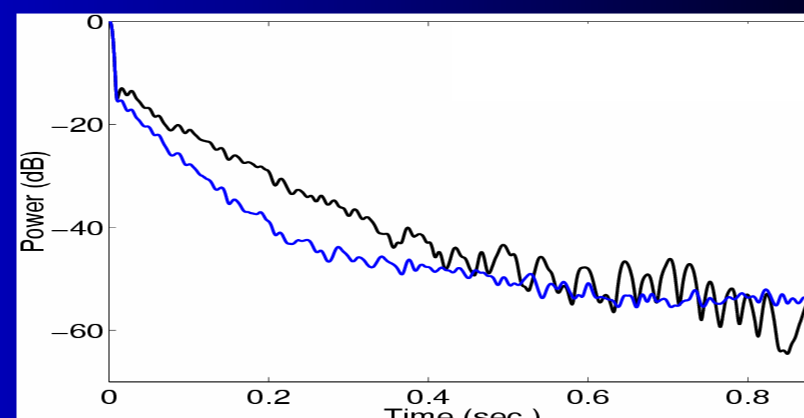
Reverberation time: 0.1 sec



Reverberation time: 0.2 sec



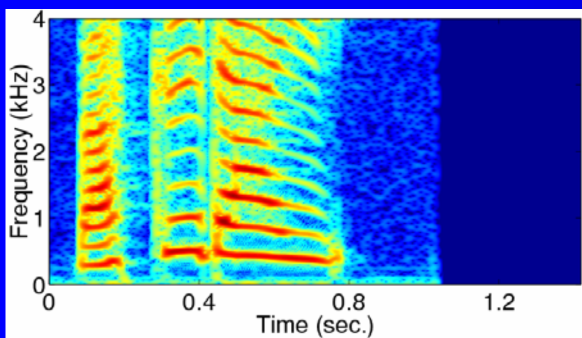
Reverberation time: 0.5 sec

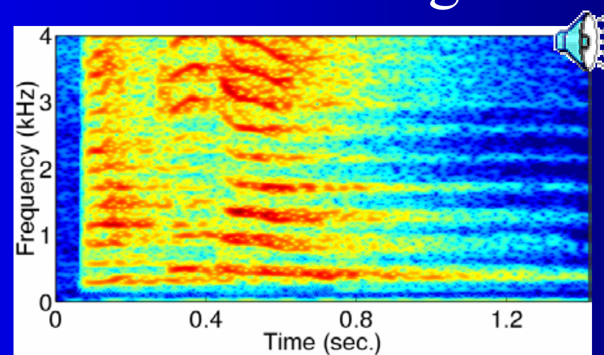


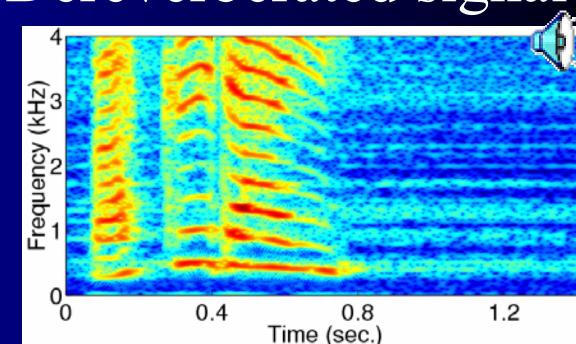
Reverberation time: 1.0 sec

Demonstration

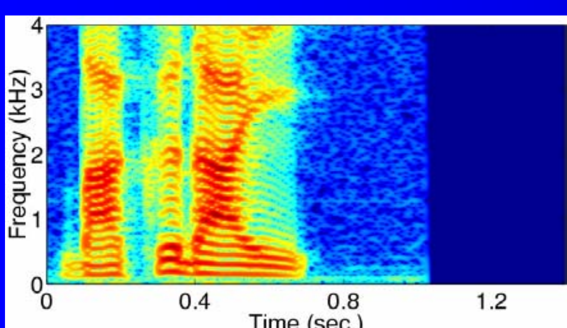
Source signal | Reverberant signal | Dereverberated signal

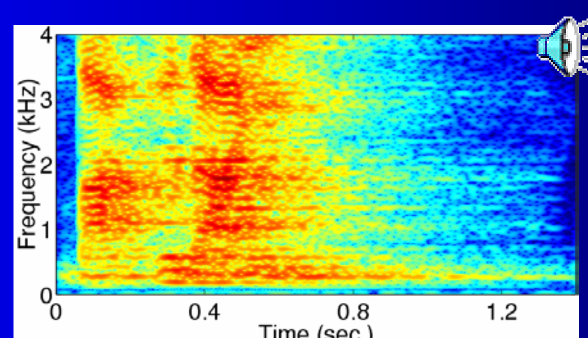


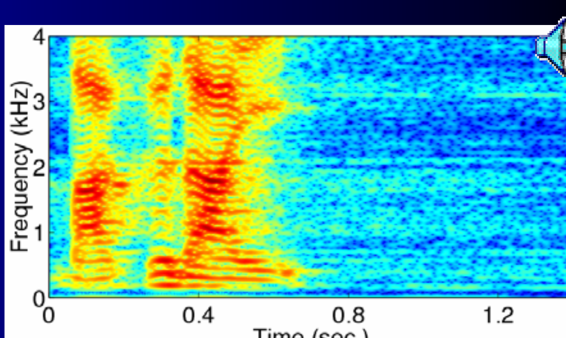




(1) Dereverberation of female voice (reverberation time: 1.0 sec)







(2) Dereverberation of male voice (reverberation time: 1.0 sec)

Coclusion

- Blind dereverberation of single channel speech signal based on harmonic structure
- Dereverberation principle is presented
- Dereverberation operator trained with 5240 words effectively reduced the reverberation
- Future work
  - Reducing the data size required for the training
  - Application to adaptive processing
  - Improving dereverberation for male speakers

WWW site for demonstration

- You can listen to our demonstration at
  - http://www.kecl.ntt.co.jp/icl/signal/nakatani/sound-demos/dm/derev-demos.html
- The above URL is also written in our paper in the Proceedings.