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
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 dereverberation
- Estimate inverse filter based on harmonics
Basic idea

Source signal

\[ X(\omega) \]

Reverberant signal

\[ Y(\omega) = H(\omega)X(\omega) \]

Approximated direct signal

\[ X'(\omega) \]

Transfer function: \( H \)

Extract a dominant sinusoid

Utilize as inverse filter

Inverse transfer function

\[ \frac{1}{H'(\omega)} = \frac{X'(\omega)}{Y(\omega)} \]
Speech modeling

Source signal

\[ X = X_h + X_n \]

- \( X_h \) : voiced (harmonics)
- \( X_n \) : unvoiced (non-harmonics)

Reverberant signal

\[ Y = HX \]

\[ Y = (D + R)X \]

- \( D \) : Direct signal component
- \( R \) : reverberant signal component

Approximated direct signal

\[ \hat{X}_h = DX_h + (\hat{R}X_h + \hat{N}) \]
Principle of dereverberation

Average inverse transfer function

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

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

Dereverberation operator

\[ O(\hat{R}) = \frac{(D + \hat{R})}{H} \]

- Reduce reverberation by multiplying \( Y \) by \( O(\hat{R}) \)

\[ O(\hat{R})Y = (D + \hat{R})X \]
Processing flow

STEP1

Reverberant signal $Y$

1. F0 estimation
2. Harmonics filtering
3. Derev. operator calculation
4. Dereverberation

STEP2

1. F0 estimation
2. Harmonics filtering
3. Derev. operator calculation
4. Dereverberation
5. Dereverberation
6. Dereverberation

Mathematical expressions:

- $F_0$
- $\hat{h}_X'$
- $\hat{O}$
- $\hat{O}'$
- $\hat{O}(\hat{R}_1)$
- $\hat{O}(\hat{R}_2)$
- $\hat{X}_{h,1}$
- $\hat{X}_{h,2}$
Harmonics filter

1. Estimate fundamental freq. $f_0(n)$ at each frame $n$

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

$$\hat{x}(t) = \sum_n \sum_k w(t - \tau_n) \hat{a}_k(n) \sin(2\pi kf_0(n)t + \hat{\phi}_k(n))$$

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

Reverberant signal $Y$ → Reduction of reverberation influence → F0 estimation → F0

F0 estimation in STEP1
- Direct harmonic sound
  - Freq. changes with time
  - Filter reducing sound components continuing at the same freq.
  - Not reduced
  - F0 estimation

F0 estimation in STEP2
- Reverberation
  - Freq. does not change
  - Derev. operator obtained in STEP1
  - $O(\hat{R}_1)Y$
  - Reduced
  - F0 estimation
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
Reverberation curves (female)

Black: original
Red: dereverberated

Reverberation time: 0.1 sec
Reverberation time: 0.2 sec
Reverberation time: 0.5 sec
Reverberation time: 1.0 sec
Reverberation curves (male)

- Black: original
- Green: dereverberated

- 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

- The above URL is also written in our paper in the Proceedings.