p1.5

ADAPTIVE DEREVERBERATION METHOD BASED ON COMPLEMENTARY WIENER FILTER AND MODULATION TRANSFER FUNCTION

Kento Ohtani†, Tatsuya Komatsu†, Takanori Nishino‡, Kazuya Takeda†
†Graduate School of Information Science, Nagoya University, Nagoya, Japan
‡Graduate School of Engineering, Mie University, Tsu, Japan

INTRODUCTION

Hands-free computer interfaces is rapidly spreading

- Speech quality are degraded by reverberation
- most of the devices have limited computational resources

Low computational dereverberation techniques are required

Previous study [K. Kondo, et al., 2012]

- Proposed a single-channel dereverberation method
- Using a complementary Wiener filter
- Need very little computational resources

Problem of previous study

- Parameters have a large impact on the performance
- Need to selecting the appropriate parameters

In this report

- Estimate the parameters of dereverberation filter
  - Using the modulation transfer function
  - Requiring very little computation

CONVENTIONAL METHOD

Relations between dereverberated speech $Z(k, l)$ and observed speech $X(k, l)$ in frequency domain

$$Z(k,l) = G(k,l)X(k,l)$$

k: frequency bin index
l: frame index

Dereverberation gain : $G(k, l)$

$$G(k,l) = \begin{cases} 1, & \dfrac{P_X(k,l)}{P_X(k)} \geq 1 \\ \dfrac{P_X(k,l)}{P_X(k)}, & \text{otherwise} \end{cases}$$

where

$$P_X(k,l) = |X(k,l)|^2$$
$$P_X(k) = E_l[|X(k,l)|^2] \qquad E_l[\cdot]\text{: expectation for } l$$

Expectation of the power spectrum

- Approximate as an exponential moving average

$$P_X(k) \approx R_2(k,l) = (1-\alpha_2)|X(k,l)|^2 + \alpha_2 R_2(k,l-1)$$
$$\alpha_2 (0 < \alpha_2 < 1)\text{: smoothing constant}$$

Instantaneous speech power

- To avoid the musical noise, using short term moving average

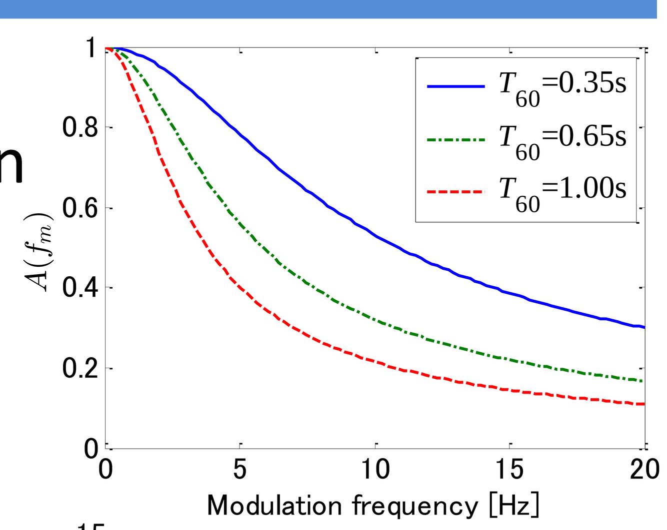
$$P_X(k,l) \approx R_1(k,l) = (1-\alpha_1)|X(k,l)|^2 + \alpha_1 R_1(k,l-1)$$
$$\alpha_1 (0 < \alpha_1 < \alpha_2)$$

PARAMETER ESTIMATION

Modulation Transfer Function (MTF)

- Represent the effect of reverberation
- MTF can be calculated as:

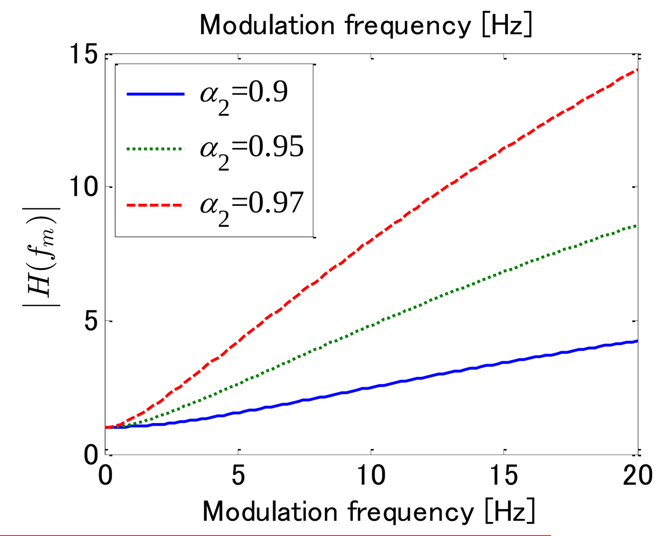
$$A(f_m) = \left\{1 + \left(2\pi f_m \frac{T_{60}}{6 \ln 10}\right)^2\right\}^{-\frac{1}{2}}$$



Amplitude response of the filter

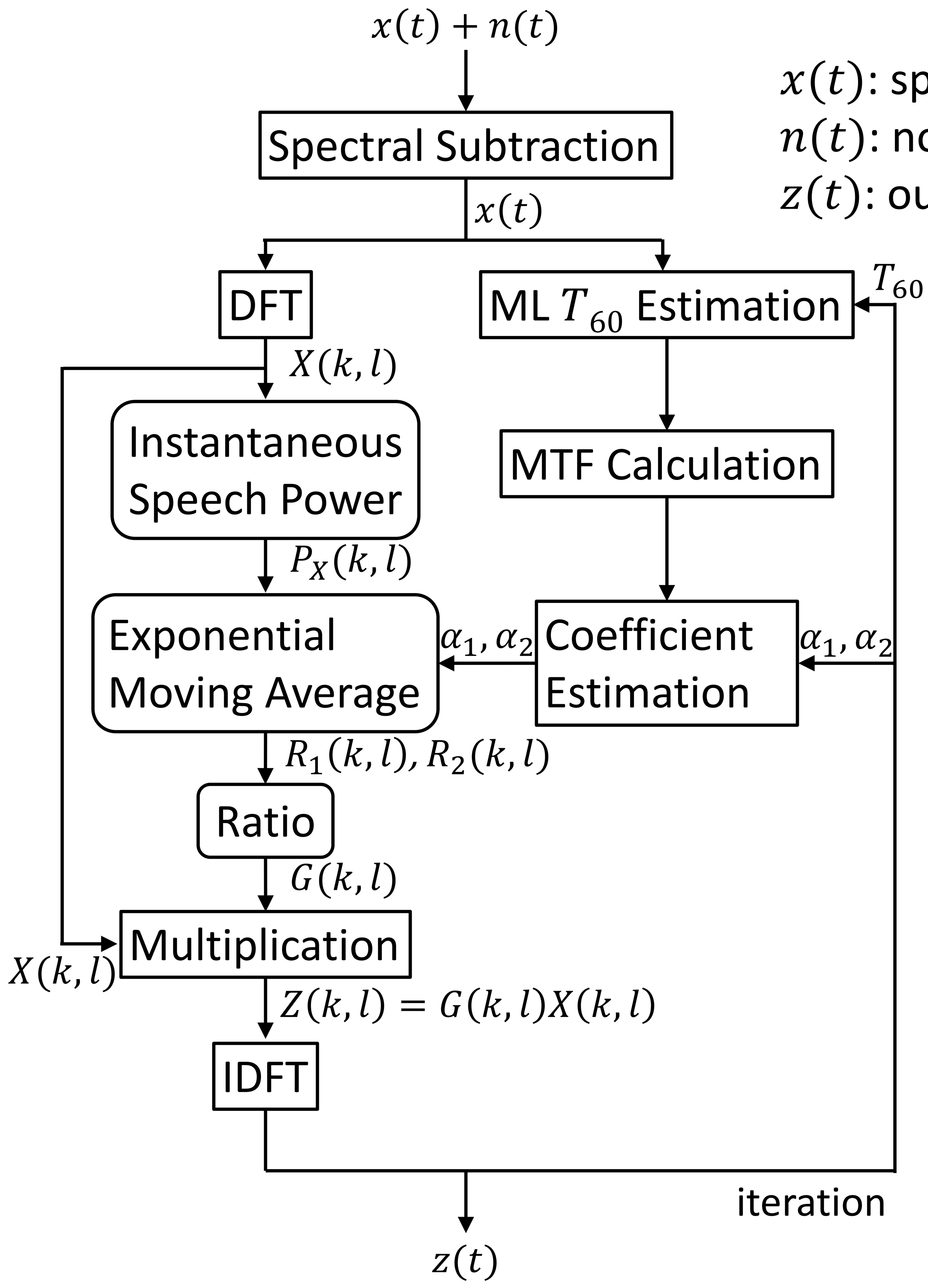
- Inverse characteristics of MTF
- This can be calculated as:

$$H(f_m) = \frac{1-\alpha_2}{1-\alpha_2}\frac{1-\alpha_2 \exp(-j2\pi f_m)}{1-\alpha_1 \exp(-j2\pi f_m)}$$



Using steepest descent method for estimation

OVERVIEW OF SINGLE ITERATION

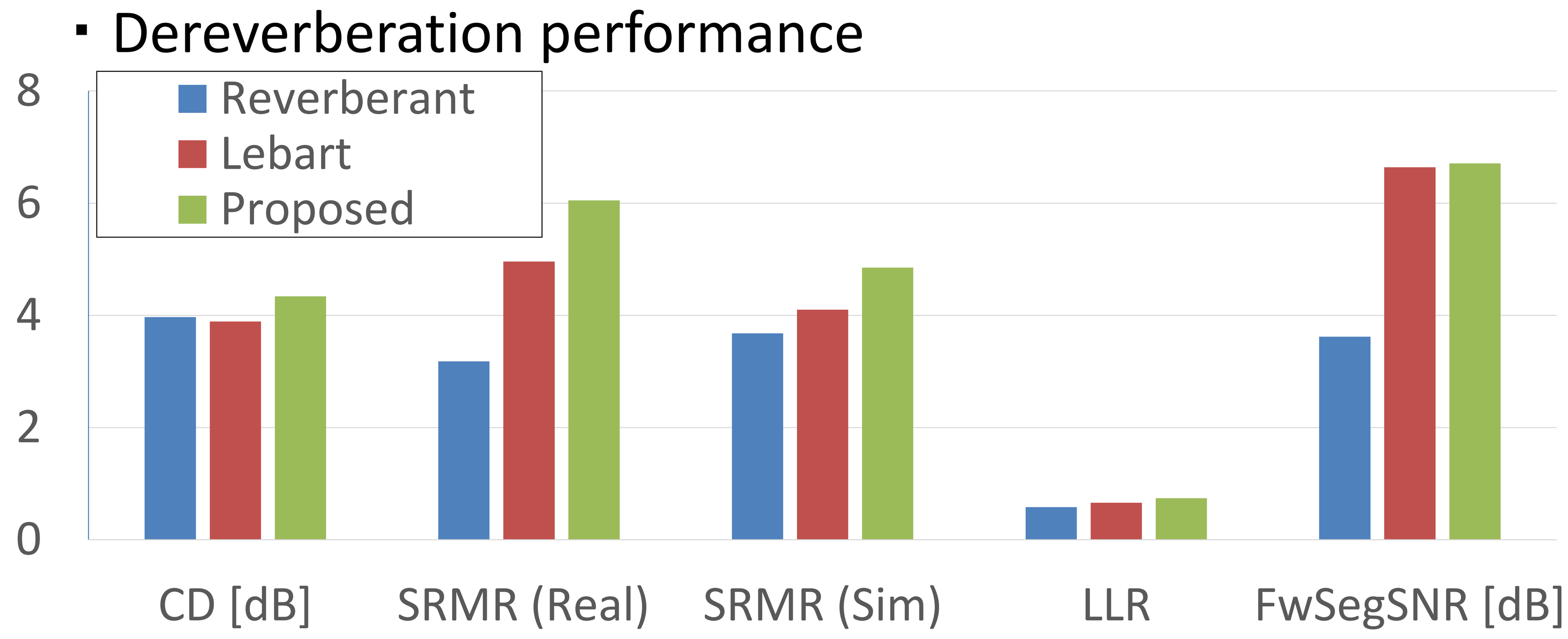


$x(t)$: speech signal
$n(t)$: noise
$z(t)$: output signal

EXPERIMENTAL RESULTS

Compared with Lebart et al., 2001

- Dereverberation performance

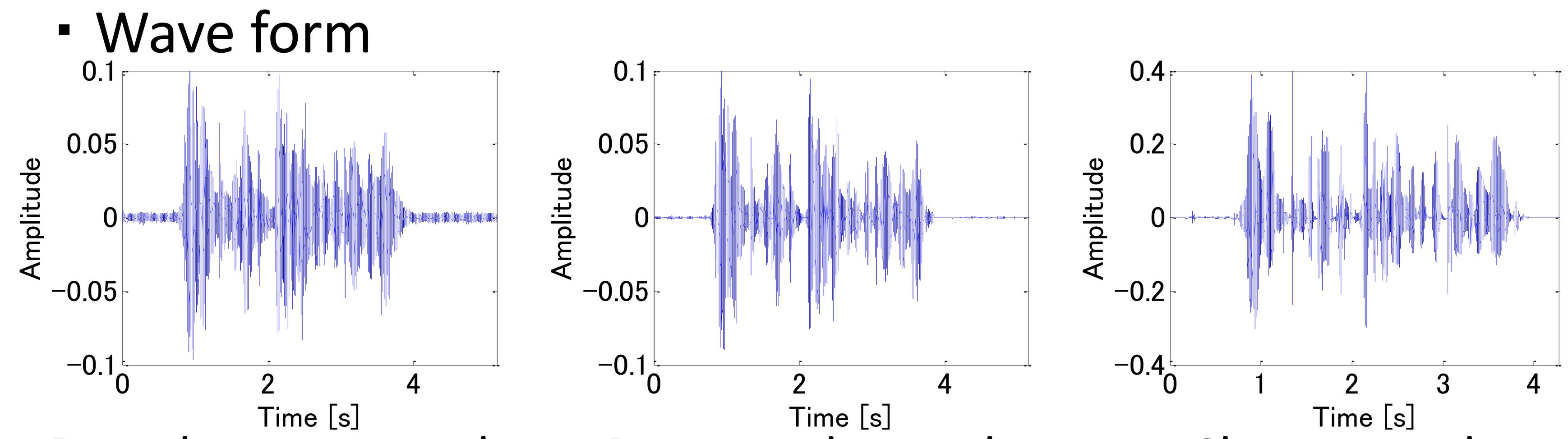


CD: Cepstrum Distance
SRMR: Speech-to-Reverberation Modulation energy Ratio
FwsegSNR: Frequency-weighted Segmental SNR
LLR: Log Likelihood Ratio

- Calculation time

Method	Proposed		Lebart et al.	
Dataset	Simulated	Real	Simulated	Real
Real time factor	0.0743	0.0707	0.3632	0.3594

- Wave form



Reverberant speech　Processed speech　Clean speech

Conclusion

Proposed a parameter estimation method

- Using MTF and amplitude response of the filter
  - They have inverse characteristics
- We can estimate the parameters properly
  - Improvement of SRMR and FwSegSNR
- Proposed method can process effectively