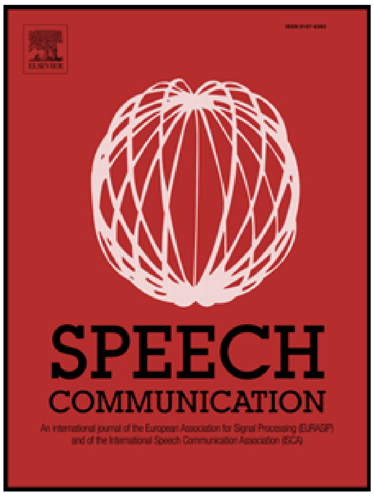
Accepted Manuscript



Model-Based Estimation of Late Reverberant Spectral Variance Using Modified Weighted Prediction Error Method

Mahdi Parchami, Wei-Ping Zhu, Benoit Champagne

PII: S0167-6393(16)30328-4
DOI: 10.1016/j.specom.2017.06.005
Reference: SPECOM 2468

To appear in: *Speech Communication*

Received date: 11 November 2016
Revised date: 5 May 2017
Accepted date: 21 June 2017

Please cite this article as: Mahdi Parchami, Wei-Ping Zhu, Benoit Champagne, Model-Based Estimation of Late Reverberant Spectral Variance Using Modified Weighted Prediction Error Method, *Speech Communication* (2017), doi: 10.1016/j.specom.2017.06.005

MODEL-BASED ESTIMATION OF LATE REVERBERANT SPECTRAL VARIANCE USING MODIFIED WEIGHTED PREDICTION ERROR METHOD

Mahdi Parchami † *Wei-Ping Zhu*† *Benoit Champagne*⋆

† Department of Electrical and Computer Engineering, Concordia University,
1455 De Maisonneuve Blvd. West, Montreal (H3G 1M8), Quebec, Canada
{m_parch,weiping}@ece.concordia.ca
⋆Department of Electrical and Computer Engineering, McGill University,
3480 University Street, Montreal (H3A 0E9), Quebec, Canada
benoit.champagne@mcgill.ca

ABSTRACT

In this paper, we propose a new approach to estimate the late reverberant spectral variance (LRSV) for speech dereverberation in the short-time Fourier transform (STFT) domain. Our approach uses a statistical model-based scheme based on a smoothing (shape) parameter and the reverberant-only component of speech for accurate estimation of the LRSV. We propose to obtain the shape parameter by using estimates of the spectral variances of the direct-path and reverberant-only components of the speech, which in turn, can be calculated by smoothing coarse estimates of these two components. Furthermore, an accurate estimate of the reverberant-only component is obtained by means of a moving average scheme over preliminary estimates of the direct-path speech terms. In order to obtain the preliminary estimates of the direct-path and reverberant speech components, we employ a modified version of the linear prediction-based dereverberation approach, namely, the weighted prediction error (WPE) method. In contrast to the original WPE method, which is implemented through batch processing of the entire speech utterance, the suggested modification is implemented for shorter processing blocks, each consisting of a number of STFT frames. This block-wise procedure allows for adaptation to moderate changes in the reverberant environment, and therefore, makes the proposed approach useful in time-varying acoustic scenarios. Performance evaluation and comparisons with previous LRSV estimation methods demonstrate the superiority of the proposed approach in both time-invariant and time-variant reverberant environments.

Index Terms— Reverberation suppression, late reverberant spectral variance (LRSV), room acoustics, short-time Fourier transform (STFT).

Abbreviations: Blind channel identification (BCI), cepstrum distance (CD), decision-directed (DD), direct-to-reverberant ratio (DRR), expectation-maximization (EM), frequency-weighted segmental SNR (FW-SNR), image source method (ISM), late reverberant spectral variance (LRSV), linear prediction coefficients (LPC), moving average (MA), minimum mean-square error (MMSE), multi-channel linear prediction (MCLP), perceptual evaluation of speech quality (PESQ), room impulse response (RIR), short-time Fourier transform (STFT), signal-to-reverberation modulation energy ratio (SRMR), signal-to-noise ratio (SNR), weighted prediction error (WPE).

Corresponding author: Mahdi Parchami, Department of Electrical and Computer Engineering, Concordia University, 1455 De Maisonneuve Blvd. W. Montreal, Quebec, Canada (H3G 1M8), email: m_parch@ece.concordia.ca

1. INTRODUCTION

Speech signals captured within an enclosure by a distant microphone are subject to reflections from the surrounding surfaces (walls, ceiling, etc.) and other objects within the environment. This phenomenon, often known as reverberation, can deteriorate the perceived quality/intelligibility of the desired speech signals, and also degrades to a large extent the performance of speech processing systems such as hearing aids, hands-free teleconferencing, source separation and localization, and automatic speech recognition systems [1, 2]. Therefore, efficient suppression of reverberation in real world acoustic environments is highly required for these applications.

During the past two decades, numerous single- and multi-microphone dereverberation methods have been developed. In the latter case, the most conventional approaches exploit beamforming techniques to coherently combine the dominant early arrivals, as in e.g., [3, 4]. However, unless a rather large number of microphones is employed, the dereverberation performance of beamforming methods is strictly limited in general [1]. Many other dereverberation approaches estimate the anechoic (clean) speech by processing observations with inverse filters that can be either calculated using the available room impulse responses (RIRs) or estimated from the reverberant observations [5, 6]. Even though perfect acoustic equalization is possible in theory if the exact RIR is known, in a realistic acoustic environment, due to the long length and ir-