Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 71 (2015) 62–73

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Objective measures for quality assessment of noise-suppressed speech

Huijun Ding [a,b], Tan Lee [c], Ing Yann Soon [d], Chai Kiat Yeo [e], Peng Dai [f,g], Guo Dan [a,b,*]

[a] Department of Biomedical Engineering, School of Medicine, Shenzhen University, China
[b] Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen, China
[c] Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong
[d] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore
[e] School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore
[f] Department of Computer Science, University of Toronto, Toronto, Canada
[g] Toronto Rehabilitation Institute, University Health Network, Toronto, Canada

Received 18 February 2014; received in revised form 26 January 2015; accepted 5 February 2015
Available online 26 February 2015

Abstract

Among all the existing objective measures, few are able to provide a clearly specific indication of speech distortion or noise reduction, which are the two key metrics to assess the performance of speech enhancement algorithms and evaluate the noise-suppressed speech quality. In this paper, new quantitative quality assessments are proposed to separately evaluate the capabilities of single channel speech enhancement algorithms in terms of maintaining the clean speech, noise reduction and overall performance. Based on these aspects, three evaluation results can be provided for any one test speech signal by analyzing the residual signal which is the difference between the clean speech and the processed speech. Several common speech enhancement algorithms are compared by these objective measures as well as subjective listening tests. High correlations between the scores of the objective measures and subjective ratings clearly show the effectiveness of the proposed evaluation methodologies on the different speech enhancement algorithms.

Keywords: Single channel speech enhancement; Speech quality assessment; Objective measure; Evaluation

1. Introduction

Single channel speech enhancement (SCSE) is an active research topic in the past decades. It aims at improving the speech quality and reducing the effects of noise through processing the noisy speech, which is the only available information for processing. A lot of milestone works have been established (Boll, 1979; McAulay and Malpass, 1980; Lim and Oppenheim, 1978; Ephraim and Malah, 1984) and developed (Loizou, 2007). Based on such abundance of work, research into evaluation methodologies follows naturally in order to evaluate and compare the different speech enhancement algorithms in scientific and efficient ways. In this paper, performance criteria are proposed for the evaluation of SCSE algorithms. The following paragraphs provide some general information on SCSE, the common evaluation methods of SCSE algorithms and the existing objective measures, followed by an overview of our proposed objective measures.

* Corresponding author at: Department of Biomedical Engineering, School of Medicine, Shenzhen University, China.
E-mail addresses: hjding@szu.edu.cn (H. Ding), tanlee@ee.cuhk.edu.hk (T. Lee), eiysoon@ntu.edu.sg (I.Y. Soon), asckyeo@ntu.edu.sg (C.K. Yeo), peng.dai@uhn.ca (P. Dai), danguo@szu.edu.cn (G. Dan).

http://dx.doi.org/10.1016/j.specom.2015.02.001
0167-6393/

SCSE problem assumes the scenario that only the noisy speech is available through one channel, and the challenge is to obtain the noise-free version from the noisy speech based on signal processing algorithms. The goal of SCSE algorithm is to remove the noise while preserving the clean speech as much as possible. However, noise reduction is always achieved at the expense of speech distortion. If the performance of SCSE algorithms on speech distortion and noise reduction can be indicated separately, a good tradeoff can be achieved by balancing both efforts, and the SCSE algorithm can be improved accordingly. Moreover, the SCSE algorithms can be selected for suitable applications to leverage the strengths based on their respective effectiveness on speech preservation and noise reduction.

Evaluation processes used to compare the different SCSE algorithms are usually applied on the resultant noise-suppressed speech signals. The straightforward evaluation method is to perform subjective listening tests which measure the speech quality by a group of human participants resulting in the Mean Opinion Score (MOS). However, such tests are time-consuming and are less reliable due to listening fatigue when a large amount of speech utterances have to be tested. Instead, objective evaluation using mathematical computations to automatically measure the speech quality is preferred by many researchers for its convenience and consistency. Therefore extensive research on objective measures has been conducted and used for many applications (Quackenbush et al., 1988; Wang et al., 1992; Vincent et al., 2006; Rix et al., 2006). It is clear that the better correlation the objective measure has with the subjective judgement, the more valid it is.

Objective measures for speech quality evaluation can be divided into two classes: parametric methods and signal-based methods. Parametric methods use key parameters obtained or estimated from the transmission system to predict the overall speech quality (Nunes et al., 2012). Signal-based methods include double-ended methods (International Telecommunications Union, 2001) and single-ended methods (International Telecommunications Union, 2004). Double-ended methods compare the observed signal with the original clean speech by measuring their numerical “distance” or “similarity”. Single-ended methods only utilize the observed signal for quality evaluation. In most SCSE simulations, the transmission characteristics are unknown and the clean speech is available. Normally the clean speech signals are randomly selected from a standard database and are corrupted by various noise. Focusing on SCSE in this paper, the double-ended methods are therefore chosen to be used and compared.

Traditional double-ended objective measurements (Klatt, 1982; Hansen and Pellom, 1998; Rix et al., 2001), typically give one evaluation score for a noise-suppressed speech to show how similar/different it is from the original clean speech. However, as stated by Loizou (2007), speech quality comprises many attributes with multi-dimensional concepts which cannot be well described by a single score. The ITU-T Recommendation P.835 (International Telecommunications Union, 2003) suggests that it is good for the noise-suppressed speech to be evaluated by the human listener in terms of three aspects: speech naturalness (speech distortion situation), called SIG, background noise intrusiveness, called BAK, and overall quality, called OVL. Three respective five-point-scale scores ranging from 1 to 5 should be given after listening. These proposals greatly cater to the needs of speech enhancement algorithm evaluations where the performance of SCSE algorithm on speech distortion and noise reduction should be individually reflected, thereby resulting in a better balance between these two efforts to achieve a good overall performance. Therefore, some pioneering work has been introduced and focuses on developing objective measures modeling various aspects of the auditory system. In Hu and Loizou (2008), the authors investigated several traditional objective measures and found that evaluation results of these objective measures have different correlations with subjective ratings in terms of SIG, BAK and OVL. Based on that, they proposed composite measures which are linear combinations of the existing objective measures to highly correlate with SIG/BAK/OVL scores and thus are able to separately measure speech distortion, noise reduction and the overall quality of the noise-suppressed speech. In Ding et al. (2010), two novel frequency-domain based objective measures are proposed to concretely analyze the filter performance and give a unique perspective in considering the speech distortion and noise reduction aspects. However, no overall quality prediction is presented.

Following this direction, three double-ended objective measures are proposed in this paper for the performance comparison of SCSE algorithms. The results of the proposed metrics, namely the computed SIG/BAK/OVL scores, are obtained by analyzing the residual signal which is the difference between clean speech and noise-suppressed speech. It should be noted that the “computed SIG/BAK/OVL scores” described in this paper does not really mean the direct estimation of subjective SIG/BAK/OVL scores. They are actually signal-based numerical calculation results, which serve the same purpose as subjective ratings, and have the same capabilities on SCSE evaluation and comparison in SIG/BAK/OVL perspectives. The block diagram of the proposed measures is shown in the Fig. 1. The computed SIG score is calculated by applying a similarity measure on the clean speech and the residual signal, both of which are also used to calculate the signal-to-noise ratio (SNR) to achieve the computed BAK score. Based on the computed SIG/BAK scores, linear regression is applied to obtain the computed OVL score. Time and frequency expressions of the proposed objective measures are derived and analyzed in this paper. In order to validate the effectiveness of the proposed objective measures, the evaluation process performed is shown in Fig. 2. Based on the listening judgements of the testing speech data, the subjective ratings are used to derive the SIG/BAK/OVL scores which are to be used as the references. The subjective ratings, as