Non-intrusive codebook-based intelligibility prediction

Charlotte Sørensen, Mathew Shaji Kavalekalam, Angeliki Xenaki, Jesper Bünsow Boldt, Mads Græsbøll Christensen

PII: S0167-6393(17)30362-X
DOI: 10.1016/j.specom.2018.06.003
Reference: SPECOM 2573

To appear in: *Speech Communication*

Received date: 25 September 2017
Revised date: 31 May 2018
Accepted date: 20 June 2018

Please cite this article as: Charlotte Sørensen, Mathew Shaji Kavalekalam, Angeliki Xenaki, Jesper Bünsow Boldt, Mads Græsbøll Christensen, Non-intrusive codebook-based intelligibility prediction, *Speech Communication* (2018), doi: 10.1016/j.specom.2018.06.003

Non-intrusive codebook-based intelligibility prediction

Charlotte Sørensen[a,b,*], Mathew Shaji Kavalekalam[b], Angeliki Xenaki[a], Jesper Bünsow Boldt[a], Mads Græsbøll Christensen[b]

[a] *GN Hearing A/S, Denmark*
[b] *Audio Analysis Lab, CREATE, Aalborg University, Denmark*

Abstract

In recent years, there has been an increasing interest in objective measures of speech intelligibility in the speech processing community. Important progress has been made in intrusive measures of intelligibility, where the Short-Time Objective Intelligibility (STOI) method has become the de facto standard. Online adaptation of signal processing in, for example, hearing aids, in accordance with the listening conditions, requires a non-intrusive measure of intelligibility. Presently, however, no good non-intrusive measures exist for noisy, nonstationary conditions. In this paper, we propose a novel, non-intrusive method for intelligiblity prediction in noisy conditions. The proposed method is based on STOI, which measures long-term correlations in the clean and degraded speech. Here, we propose to estimate the clean speech using a codebook-based approach that jointly models the speech and noisy spectra, parametrized by auto-regressive parameters, using pre-trained codebooks of both speech and noise. In experiments, the proposed method is demonstrated to be capable of accurately predicting the intelligibility scores obtained with STOI from oracle information. Moreover, the results are validated in listening tests that confirm that the proposed method can estimate intelligibility from noisy speech over a range of signal-to-noise ratios.

Keywords: Hearing aids, non-intrusive, speech intelligibility prediction, STOI

1. Introduction

Human interaction depends on communication where speech has a central role. Inability to understand speech, e.g., due to hearing impairment, noisy background, or distortion in communication systems, can lead to ineffective communication and social isolation, and the development of speech enhancement methods[1, 2] is, therefore, a key concern in many applications. These include challenging applications such as hearing aids [3], telecommunication systems [4, 5], and architectural acoustics [6]. To assess the listening conditions in which speech processing would be beneficial, but also to evaluate the speech processing algorithms as such, a speech intelligibility measure is required [3, 5, 7].

A natural way of assessing the intelligibility of a degraded, i.e., processed, distorted or noisy speech signal is by performing subjective listening tests. Subjective speech intelligibility scores gives the percentage of correctly identified information from a degraded speech signal. However, subjective speech intelligibility experiments are time-consuming, expensive and cannot be used for real-time applications. Hence, there is a great interest in developing objective measures for speech intelligibility prediction. As opposed to subjective listening tests, objective intelligibility prediction algorithms are faster, cheaper and can be used for real-time processing.

The Articulation Index (AI) [8, 9] and the Speech Intelligibility Index (SII) [10] are some of the earliest metrics for prediction of speech intelligibility scores. The AI and SII use the signal-to-noise ratio (SNR) of speech excerpts in several frequency bands to estimate the intelligibility, hence they require that both the clean speech signal and the noise are available and uncorrelated as well as the noise to be stationary. The Extended SII (ESII) [11] and the Coherence SII (CSII) [12], are variants of SII which account for fluctuating

*Corresponding author.
Email addresses: csoerensen@gnresound.com, cs@create.aau.dk (Charlotte Sørensen), msk@create.aau.dk (Mathew Shaji Kavalekalam), axenaki@gnresound.com (Angeliki Xenaki), jboldt@gnresound.com (Jesper Bünsow Boldt), mgc@create.aau.dk (Mads Græsbøll Christensen)

[1]This work was supported by the Innovation Fund Denmark, Grant No. 99-2014-1.