Iterative Refinement of Amplitude and Phase in Single-Channel Speech Enhancement

Pejman Mowlae†, Mario Kaoru Watanabe† and Rahim Saeidi‡
† Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria
‡ Speech and Image Processing Unit, School of Computing, University of Eastern Finland, Finland
pejman.mowlae@tugraz.at mario.watanabe@student.tugraz.at rahim.saeidi@uef.fi

Abstract

While the state-of-the-art speech enhancement methods are focused on the modification of the noisy spectral amplitude, our recent findings demonstrate positive impact of incorporating the speech phase spectrum in speech enhancement. In this show and tell proposal, we demonstrate the recent progress towards utilizing the phase information in closed-loop iterative manner leading to the joint enhancement of amplitude and phase spectra. In this show and tell, we provide a demo where a recorded speech corrupted with noise is enhanced by an iterative refinement of amplitude and phase. The improvement in speech enhancement in various signal-to-noise ratios is justified by improved perceived speech quality as well as not degrading the speech intelligibility performance.

Index Terms: Speech enhancement, phase-aware amplitude estimation, iterative closed-loop speech enhancement.

1. Introduction

In various real world applications including mobile communication and hearing aids, speech acquisition needs to be robust to noise. In particular, to maintain a robust performance against the interfering noise, a speech enhancement technique is necessary. With the advent of more noisy scenarios occurring in real world applications experienced in the everyday life speech acquisition, new effective solutions are required to further push the limited performance of the conventional enhancement techniques. In this regard, the speech phase spectrum is one of the missing aspects considered as a controversial topic within the past three decades. The earlier studies e.g. [1] pointed on the unimportance of the phase in speech while more recent studies demonstrate its positive impact in speech enhancement [2–6] or other speech applications studied in [7].

Inspired by the aforementioned findings on the importance of phase in speech processing, in this show and tell we focus on demonstrating phase-aware speech enhancement recently proposed in [2] where the amplitude and phase of the underlying speech signal are jointly enhanced. According to Figure 2, at each iteration a new estimation of the speech phase spectrum is obtained by using the phase estimator of [8] and is passed to a phase-aware spectral estimator [9], and is further associated with the estimated phase to produce a consistent complex spectrum using one iteration of the so-called Griffin and Lim algorithm [10]. The so-obtained enhanced speech signal is shown to have better perceived quality and speech intelligibility compared to state-of-the-art techniques (for a more detailed review see [7]). This is an interesting observation since according to the previous findings reported in [11] the conventional speech enhancement methods improve the perceived quality at the expense of reducing the intelligibility of the noisy speech.

2. Concept From the Technical Viewpoint

We proposed an iterative closed-loop approach in [2], where in each iteration we estimate phase spectrum, next use this phase to estimate the amplitude spectrum and eventually make the complex spectrum more consistent. A demonstration of the performance for our proposed method in [2] is shown in Figure 4.

3. Proposed Full system

Innovative Aspects: In this show and tell, we demonstrate the impact of incorporating the estimated phase spectrum towards improving the speech spectral amplitude in a closed-loop configuration as shown in Figure 2. As the starting point, an initial estimation for the amplitude spectrum is provided by one of the many conventional methods (see section 2.2). Then the clean speech phase spectrum is estimated using the geometry and group delay constraints described in [8]. The so-obtained estimated phase is passed to a phase-aware spectral amplitude estimator proposed in [9]. The estimated phase and phase-aware spectra are glued together followed by an iteration of synthesis-analysis denoted by STFT(STFT−1(·)). By doing so, a consistent complex spectrum is obtained (see e.g. [10]). As stopping criterion the inconsistency level at each iteration is checked and compared to the previous iteration. The iteration is terminated once the difference is lower than a pre-defined threshold and the final enhanced signal is produced using overlap and add.

Results: The performance of the iterative closed-loop phase-aware method is evaluated in terms of perceived speech quality and speech intelligibility, measured by PESQ [12] and SNR俣 [13], respectively, both shown to have high correlation with subjective listening tests [14]. Figure 4 shows the
of the genders of the speakers. Figure 2 shows the snapshot among white and babble together with different possibilities ranging in 0 to 15 decibels and different noise scenarios. The user can select between different signal-to-noise ratios scheduled for presentation at show and tell session. In the offline mode, a noise corrupted speech signal is presented for evaluation during speech enhancement experiments.

The joint enhancement of phase and amplitude spectra removes the limits on the achievable performance of the conventional methods where noisy phase is used for reconstruction, or the amplitude estimator assuming the independent assumption on both the limits on the achievable performance by the conventional method and is able to recover most of harmonic structure in speech lost due to addition of noise. The iterative closed-loop configuration improves the perceived speech quality while maintaining the speech intelligibility reduction lower than other methods. This trade-off is quite interesting since, as demonstrated in [11], the speech enhancement methods are already known to reduce the speech intelligibility of the noisy signal. The joint enhancement of phase and amplitude spectra removes the limits on the achievable performance by the conventional methods where noisy phase is used for reconstruction or the amplitude estimator assuming the independent assumption on phase and amplitude spectra.

**Presentation Format**: Our proposed demonstration will run on a Laptop and the listeners are equipped with headphones during the session. MATLAB Graphical User Interface (GUI) is used as the user interface. In order to simultaneously serve several listeners participating in our show and tell demo, we designed a filter box which output up to 8 channels. A sample video snapshot for the show and tell demonstration is available here http://www2.spac.tugraz.at/people/pmowlae/Video.wmv which overviews the contents scheduled for presentation at show and tell session. In the offline mode, a noise corrupted speech signal is presented and the user can select between different signal-to-noise ratios ranging in 0 to 15 decibels and different noise scenarios among white and babble together with different possibilities of the genders of the speakers. Figure 2 shows the snapshot of the running GUI for a female speaker speech signal corrupted with babble noise at SNR = 0 (dB). The top and bottom panels display the spectrogram and the group delay plots for the selected clip. Audio wave files are downloadable at http://www.spac.tugraz.at/st2014.

4. References