1. Introduction

Speech produced under real conditions (not a recording studio, nor a quiet room) is not always equally intelligible due to the presence of background noise. This noise may mask part of the speech signal such that not all speech information is available to the listener. The ability to detect speech in noise plays a significant role in our communication with others. In this work we suggest the use of a non-parametric way to improve the intelligibility of speech under adverse noisy conditions by modifying the speech signal accordingly.

2. Method

The suggested system contains two subsystems: (i) Spectral Shaping (SS) and (ii) Dynamic Range Compressor.

The goal of Spectral Shaping is to increase the “crisp” and “clean” quality of the speech signal, and therefore improve the intelligibility of speech even in clear (not-noisy) conditions. For this, both adaptive and fixed spectral shaping operators are used. The adaptive spectral shaping takes into account the probability of voicing given a speech frame, while the fixed spectral shaping is independent of the probability of voicing. The adaptive spectral shaping consists of (i) adaptive sharpening where the formant information is enhanced, and (ii) an adaptive pre-emphasis filter. The adaptive (to the probability of voicing) characteristic of the suggested system is important for not introducing artifacts in the processed signal especially in fricatives, silence or other “quiet” areas of speech. The purpose of the fixed (non-adaptive) spectral shaping is to protect the speech signal from low-pass operations during the reproduction of the speech signal.

The output of the Spectral Shaping system is the input to the Dynamic Range Compressor (DRC). DRC has a dynamic and a static stage. During the dynamic stage, the envelope of the signal is dynamically compressed with 2ms release time constant and almost instantaneous attack time constant. The signal envelope is based on the Hilbert transform and a moving average operator with order determined by the average pitch of speaker’s gender. After the dynamic compression of the signal envelope, a static amplitude compression is applied. During the static amplitude compression, the 0 dB reference level is a key element in forming the Input/Output Envelope Characteristics (IOEC). For the current system this was set to 0.3 the peak of the signal.

The whole system is based on a frame-by-frame analysis and synthesis. In each frame the magnitude spectrum is computed using FFT and then manipulated in the way mentioned above. Overlap and add is then used to reconstruct the modified signal. The whole process is very fast and can run in real time.

3. Results

For testing the system, we used the first 20 Harvard sentences and two types of noise: Speech Shaped Noise (SSN) at SNR: -9 dB, -4 dB and 1 dB, and Competing Speaker noise (CS) at SNR: -21 dB, -14 dB, -7 dB. For evaluation and comparison purposes, the extended SII suggested in [1] and the frequency dependent SNR recovery system suggested in [2], were implemented. Fig. 1 shows the results in terms of SII for the original signal without modifications (Orig), the suggested system (SS-DRC) and the system presented in [2] (referred to as SNR-R). Performance of sub-systems (Spectral Shaping (SS), Dynamic Range Compression (DRC)) is also shown. The final system combines SS and DRC in a cascade form. Overall, the suggested system (SS-DRC) outperforms SNR-R for all SNR levels and for both types of noise. All modified signals (either modified by SNR-R or SS-DRC) report better SII score than the non-modified signals.

4. References
