Simple Spectral Techniques to Enhance the Intelligibility of Speech using a Harmonic Model

Daniel Erro¹, Yannis Stylianou²,¹, Eva Navas¹ and Inma Hernaez¹

¹Aholab Signal Processing Laboratory, University of the Basque Country (UPV/EHU), Bilbao, Spain
²Institute of Computer Science, FORTH, and Multimedia Informatics Lab, CSD, UoC, Greece
{derro, eva, inma}@aholab.ehu.es, yannis@csd.uoc.gr

Abstract
We have designed a tool to increase the intelligibility of speech by manipulating the parameters of a harmonic speech model. The system performs the transformation in two steps. In the first step, it modifies the spectral slope, which is closely related to the vocal effort. In the second step, it amplifies low-energy parts of the signal using dynamic range compression techniques. Such a system has two main advantages: its simplicity and the fact that it can be easily integrated into the synthesis engine of a speech synthesizer trained from Mel-cepstral coefficients.

1. Introduction
Speech synthesizers are usually trained from clean speech databases recorded by professional speakers in silent environments. Consequently, when synthetic speech is played in noisy conditions it is often hard for listeners to understand the message. For speech synthesizers to be practical in different contexts, it is desirable to have some control over the voice characteristics that play a crucial role in intelligibility.

There are basically two ways of modifying the synthetic speech to make it more intelligible in noisy conditions: (i) acquiring the right databases and then using statistical mapping techniques [1]-[4]; (ii) using expert knowledge and signal processing techniques to enhance the output of the synthesizer [3][5]. This work follows the latter approach, which is advantageous in the sense that it avoids recording new databases and retraining the underlying models.

In this paper we show that the parameters of the harmonic model can be modified to increase speech intelligibility in noise. Since the harmonic model is applicable to synthesizing high-quality speech from a statistically generated Mel-cepstral vector sequence [6], the proposed transformation can be easily integrated into the synthesizer, which makes it able to modify its output according to the environmental noise at the expense of an almost negligible increment of the computational load.

2. Brief Description
The harmonic model assumes that locally stationary speech signal segments can be decomposed into a series of harmonically related sinusoids represented by their frequencies, amplitudes and phases. The proposed system operates entirely on the parameters of this model. It consists of two transformation steps to be applied in cascade.

Lombard speech is characterized by a higher vocal effort, which has an impact on the spectral tilt. We have studied a simple spectral transformation that modifies the amplitudes by a constant slope measured in dB/decade and then renormalizes their energy at frame-level.

Under the hypothesis that nonstationary portions of the speech signal (plosives, for instance) play a decisive role in intelligibility, the second transformation studied in this work aims at amplifying low-energy frames, which are likely to contain most of these meaningful portions. We have explored modification procedures inspired by Dynamic Range Compression (DRC) techniques, which can also be implemented in the harmonic amplitude domain.

3. Results
Objective tests based on our implementation of the extended Speech Intelligibility Index (SII) [7] indicate that the intelligibility of natural signals in speech-shaped noise improves significantly after manipulation. The two steps of the transformation contribute cumulatively to this improvement, each one increasing the SII score by approximately 0.1. Formal subjective evaluations carried out by 78 native listeners confirm that the enhancement system reduces the median word error rate at sentence level by 30-50% for -9dB ≤ SNR ≤ 1dB.

4. Acknowledgements
This work has been partially supported by the Spanish Ministry of Science and Innovation (Buceador Project, TEC2009-14094-C04-02) and the Basque Government (ZURE_TTS Project).

5. References