



Speech enhancement and psychoacoustics

Asmaa Amehraye, Dominique Pastor

► To cite this version:

Asmaa Amehraye, Dominique Pastor. Speech enhancement and psychoacoustics. European Mathematical Psychology group : 37ème meeting, Sep 2006, Brest, France. hal-02136736

HAL Id: hal-02136736

<https://hal.science/hal-02136736>

Submitted on 22 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Speech enhancement and psychoacoustics

Asmaa Amehraye ^{*}, Dominique Pastor [†]

Abstract

The purpose of this presentation is to describe why and how psychoacoustic models are used to design speech enhancement systems capable of improving speech intelligibility in presence of noise (for mobile communication applications, for instance) or capable of denoising sufficiently well noisy speech signals so as to improve the recognition rate of some automatic speech recognizer (for instance, robot monitoring in noisy environment, hand-free applications on board of vehicles, military fastjets and helicopters).

To begin with, standard methods aimed at denoising speech signals are performed in the spectral domain without taking into account the perceptual characteristics of the speech signal to enhance. They succeed in improving the Signal to Noise Ratio (SNR) but return annoying and unpleasant residual noise known as musical noise.

In the last few decades, psychoacoustic models have then attracted a great deal of interest. The objective is to improve the perceptual quality of the enhanced speech signal. The psychoacoustic model is used to control the enhancement process in order to find the best trade-off between noise reduction, residual noise and speech distortion. The masking phenomenon is the main human auditory system property which is used to design perceptually motivated speech enhancement systems.

The masking phenomenon occurs when the maskee is made inaudible by a stronger signal that occurs simultaneously. We model this phenomenon by a masking threshold calculation that depends on the signal to denoise ([3]). Any signal below this threshold is inaudible. The main idea is then

^{*}A. Amehraye is with the GSCM-Faculté des Sciences de Rabat, Morocco and also with Ecole Nationale Supérieure des Télécommunications de Bretagne asmaa.amehraye@enst-bretagne.fr

[†]D. Pastor is with Ecole Nationale Supérieure des Télécommunications de Bretagne, Technopôle de Brest-Iroise, CS 83818, 28238 Brest, France dominique.pastor@enst-bretagne.fr

to mask partially or totally the distortion. We will present and comment several filters based on this principle ([2], [1]).

Psychoacoustic models are also very popular to assess the quality of speech after coding or enhancement. By using objective tests based on these models, we avoid subjective tests that are very costly in time and resources. The most used objective criteria based on psychoacoustic models are the MBSD (Modified Bark Spectral Distortion) and the PESQ (Perceptual Evaluation of Speech Quality) [4], [5], [6]. We will present how these objective tests are used and commented in relation with standard and perceptually motivated denoising techniques.

References

- [1] A. Amehraye, D. Pastor, S. Ben Jebara, On the Application of Recent Results in Statistical Decision and Estimation Theory to Perceptual Filtering of Noisy Speech Signals, *Proceedings of the second IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing*, ISCCSP'06, Marrakech, Morocco, 2006.
- [2] Y. Hu and P. Loizou, "Incorporating a psychoacoustic model in frequency domain speech enhancement", *IEEE Signal Processing Letters*, 11(2), 2004, pp. 270-273.
- [3] Johnston, J. D, "Transform coding of audio signals using perceptual noise criteria", *IEEE Jour. Selected Areas Commun.*, vol. 6, no. 2, 1988, pp. 314-323.
- [4] S. Quackenbush, T. Barnwell, and M. Clements, "Objective Measures of Speech Quality". *Englewood Cliffs, NJ: Prentice-Hall*, 1988.
- [5] S.Wang, A. Sekey and A.Gersho, "An objective measure for predicting subjective quality of speech coders", *IEEE Journal on Selected Areas in Communications* , vol. SAC-10, 1992, pp. 819-829.
- [6] W. Yang, M. Dixon and R Yantorno, "Modified bark spectral distortion measure which uses noise masking threshold", *IEEE Speech coding Workshop*, Pocono Manor, 1997, pp. 55-56.