Perceptual consequences of amplitude perturbations in the wavelet coding of speech

Nicolle H. van Schijndel, Tammo Houtgast and Joost M. Festen

Department of Otolaryngology, University Hospital VU, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands, NH.vanSchijndel@azu.nl

Abstract: Wavelet coding of sound is a frequency-time representation which can, to a certain extent, mirror the properties of auditory coding at a peripheral level, provided that we use the proper so-called mother (prototype) wavelet. Previous experiments based on intensity discrimination (1) suggested a Gaussian-shaped wavelet with a bandwidth of 1/4 octave. Based on this function, a wavelet coding and reconstruction algorithm was developed. In an attempt at studying the effect of imperfect auditory amplitude coding, the amplitude of the wavelet coefficients is manipulated before the stage of reconstruction and the perceptual effect of these amplitude perturbations on speech perception is studied. Random amplitude perturbations within a few dB are detectable to the listener. Speech intelligibility decreases almost linearly with supra-threshold amplitude perturbation. These results will serve as a reference set for experiments with hearing-impaired listeners.

INTRODUCTION

In analyzing nonstationary sounds, the auditory system carries out a frequency-time analysis. Since a basic entity of the auditory process is the critical band, emulating the hearing process may naturally lead to a wavelet analysis, which is also characterized by a constant relative bandwidth, like the critical band. Experiments of Van Schijndel et al. (1) suggested a Gaussian-shaped mother wavelet with a bandwidth of 1/4 octave. So, in a first approximation, the auditory system was modeled as a wavelet analysis using a 1/4 octave Gaussian wavelet. The main point of interest in this study is the perceptual effect of the amplitude perturbation of the wavelet coding coefficients. Basic questions are (1) what degree of amplitude perturbation is detectable to the listener, and (2) how is speech intelligibility affected by supra-threshold amplitude perturbation. The results may provide insight into the acuity of a listener's auditory amplitude coding and into the practical consequences in terms of hearing performance when this acuity is (artificially) reduced.

METHOD

The speech stimuli are wavelet coded using a Gaussian wavelet with an effective bandwidth of 1/4 octave. The frequency-time plane is discretized with a resolution of one wavelet per three periods along the time axis and eight wavelets per octave along the frequency axis. Before the stage of reconstruction the amplitude levels of all wavelet coding coefficients were perturbed with a random value chosen from a uniform distribution ranging from -Pmax dB to +Pmax dB.

Ten normal-hearing listeners participated in the experiment. The stimuli were presented in the middle of the dynamic range of the listeners. In experiment 1 the amplitude perturbation detection threshold was estimated for monosyllabic words. To estimate roughly the influence of the particular frequency range stimulated, ninety-six CVC-words (2) were bandpass filtered (250-4000 Hz, 250-1000 Hz and 1000-4000 Hz) and perturbed. The words were presented randomly in a 3I-3AFC, 2-down 1-up adaptive procedure, leading to a 70.7%-correct score. In experiment 2 the influence of the amplitude perturbation on speech intelligibility for sentences was investigated (for a 250-4000 Hz bandpass filtered condition). The SRT in noise was determined according to Plomp and Mimpen (3) in which the analysis-perturbation-reconstruction scheme was applied to the speech-plus-noise stimulus.

RESULTS

The individual perturbation detection thresholds for 10 listeners are presented in Figure 1a. The broadband condition (250-4000 Hz) resulted in an average detection threshold of 4.9 dB. In the low-frequency condition (250-1000 Hz) the threshold was 4.9 dB. In the high-frequency condition (1000-4000 Hz) the threshold equals 6.9 dB. The high-frequency threshold is significantly higher than the low-frequency threshold. No significant