Reducing Sibilants in Recorded Speech using Psychoacoustic Models

Markus Sapp, Martin Wolters* and Jörg Becker

Institute of Communication Engineering (IENT), Aachen University of Technology (RWTH), 52056 Aachen, Germany (* now with Telos Systems, Cleveland, USA)

Abstract: Sibilants are a known problem in speech recording. Even though they can often be decreased by a different placement of the microphones, there still is the necessity for methods that reduce these artefacts. Therefore new investigations on sibilants in German, English, Spanish, and French have been made to describe their properties in time- and frequency-domain. To find an adaptive algorithm which considers these properties, a reliable method is needed for detecting and classifying the different kinds of sibilants. It is shown that the psychoacoustic unit 'sharpness' is well correlated with the appearance of these sounds and that the critical-band intensities - an intermediate step in calculating sharpness - can be utilised to get information about the spectral properties of each sibilant. An algorithm is presented which employs sharpness to detect sibilants and reduces them using variable filters.

INTRODUCTION

The main objective in speech recording is to produce a highly realistic image of the original acoustic event. During the recording process signal distortions and artefacts might be introduced. While linear distortions can easily be compensated using equalizers, the elimination of non-linear distortions and other artefacts is more complicated. One frequently occurring problem, partly depending on the position of the microphone, are sibilants. In the past, various systems for reducing these artefacts were developed, known as DeEsser. Most of them work in the analog domain, using dynamic compression or spectral subtraction. The rate of compression, respectively the weighting of subtraction, is controlled by the percentage of high-frequency energy in the input signal. In this paper, a new algorithm for digital implementation is proposed which uses psychacoustic signal analysis for detecting and reducing sibilants.

DETECTION OF SIBILANTS

Some consonants tend to be sensed as sibilants when recorded. Mainly these are fricatives, affricates or plosives. In some phonetic books the term sibilant is used for stridents, a subclass of fricatives with high noise energy (1), but in this paper the term will be used in the more general sense.

For a reliable detection of sibilants, knowledge about striking properties of these artefacts is needed. Therefore a listening test was performed with four experts. The test-material consisted of sentences from four languages (German, Spanish, French and English) spoken by native speakers and recorded simultaneously with two types of microphones (Shure SM58 and Neumann TLM170) in an anechoic chamber. The test persons were asked to mark all passages where disturbing sibilants occurred and rate the strength of sibilance with values from 0.5 (hardly disturbing) to 4 (very disturbing). From this listening test 141 locations in the test sentences resulted where at least three of the four test persons detected a disturbing sibilant. These passages were then further analyzed.

The semantic relation of the terms sibilant and sharp and the fact that high noise energies are characteristic for sibilants lead to a closer look at the psychacoustic unit sharpness (2). Sharpness is a weighted first moment of specific loudness which in turn is an intermediate step in calculating loudness (3). For all 141 sibilants detected by the test persons and for all 50 complete test sentences, the sharpness was calculated for blocks of 256 samples. From this data the mean and standard deviation of sharpness was derived for each signal. Figure 1 shows deviation versus mean of sharpness for all 191 signals. It can easily be seen that sibilants can safely be distinguished from non-sibilants. The larger standard deviations for the whole test sentences result from the included sibilants. Resulting from these studies, the presented algorithm detects a sibilant when the calculated sharpness exceeds 1.2 acum. The end of a sibilant can be detected by sharpness falling significantly below this value.
REDUCTION OF SIBILANTS

To reduce a detected sibilant, the disturbing part of the input signal is extracted using cascaded linear-phase high-pass and low-pass filters, scaled by a weighting factor and subtracted from the correctly delayed input signal. The critical-band intensities, an intermediate step in calculating sharpness, represent aurally adequate information about the spectral properties of the sibilant. The frequencies nearest to the maximum where the critical-band intensity falls below a certain percentage of the maximum, are taken as cutoff-frequencies for the high-pass and low-pass filters. Percentages between 20% and 40% lead to good results, with smaller values increasing the bandwidth of the reduction. The weighting factor specifies the strength of reduction and is calculated from sharpness and bandwidth. It is smoothly faded in and out for each sibilant.

Since there were no research results available concerning the temporal development of the spectra of sibilants, the critical-band intensities of the 141 sibilants detected in the listening test were analyzed in blocks of 256 samples with a frequency resolution of 0.5 bark. For each block the frequency of the maximum and the nearest frequencies where the critical-band intensity falls below 20%, respectively 40%, of the maximum were determined. In figure 2 these values are depicted versus time for one representative example. It can be seen that the spectral properties are constant for the duration of the sibilant, so that the cutoff-frequencies need not be updated until a new sibilant is detected. Sibilants directly following each other are separated by a minimum in signal energy, so new reduction parameters can be calculated after occurrence of such a minimum.

CONCLUSIONS

An algorithm for detecting and reducing sibilants was presented. It is based on the determination of critical-band intensities and sharpness of the input signal. It was shown that sharpness is a reliable indicator for the occurrence of sibilants and that the spectral properties are almost constant for the duration of each sibilant. First tests showed good reduction results, but there still remains the necessity of extensive listening tests for further parameter optimization. For an effective real-time implementation of the presented algorithm some modifications and simplifications are useful (4).

REFERENCES