Estimation of Noise Spectrum and its Application to SNR-Estimation and Speech Enhancement

H.Günter Hirsch

Technical Report TR-93-012

International Computer Science Institute,
Berkeley, California, USA

Estimation of Noise Spectra and its Application to SNR-Estimation and Speech Enhancement

H. Günter Hirsch

Contents

1. Introduction

At the time of this writing, some experiments in robust speech recognition had already been done at ICSI.
The original goal of this work was improved recognition of speech recorded with different microphones and transmitted over channels with different frequency characteristics. One practical application of this is the recognition of speech recorded via telephone lines where you have microphones and channels with different transmission characteristics. It could be shown that the recognition rates can be improved when introducing a high-pass filtering of the logarithmic spectral envelopes in subbands /1/.
This idea is based on the fact that a frequency characteristic corresponds to a multiplication of the speech spectrum with the frequency response of the transmission channel. The result would be a constant additive component in the logarithmic spectral envelopes in subbands (assuming a nearly constant transmission characteristic). Because of this, a high-pass filtering leads to a suppression of these constant components.

Another aspect is the superposition of noise in many applications of speech recognizers in real environments, e.g. voice dialing in a car or serving any kind of machines on the street or in workshops. This noise would result in a nearly constant additive component to the magnitude spectral envelopes in subbands (assuming a nearly stationary noise) . It could be shown that recognition rates can be improved by high-pass filtering the magnitude spectral envelopes /2/.

Additive noise as well as a certain frequency characteristic are present in many real situations. One way to handle both effects could be to use a combination of processing in the magnitude as well as in the logarithmic spectral domain. Another possibility could be a processing anywhere between the magnitude and the logarithmic domain dependent of the amount of noise in the specific situation. This would presuppose an estimation of the signal-to-noise ratio (SNR).

Looking at the first possibility several processing techniques are well known to reduce the noise in the magnitude spectral domain. One could be the already mentioned high-pass filtering. A disadvantage of this method is the suppression of certain spectral features in speech segments. Introducing a high-pass filter with a total suppression of the DC component, not only the constant noise components are suppressed but also the constant component of the speech. Because of this just the spectral features of the phonemes with less energy are reduced in the case of a preceding phoneme with higher energy and spectral components in the same subbands.

One solution could be a kind of nonlinear filtering with the goal of preserving the spectral features of the phonemes with less energy on one hand but suppressing the noise components on the other hand. Another method to reduce the noise is the well known spectral subtraction technique /3/,/4/. This technique is based on the estimation of the noise spectrum during speech pauses and an adaptive filtering with the estimated noise spectrum. A major disadvantage is the necessity of the detection of speech pauses to estimate the noise spectrum. This is a very difficult and ultimately unsolved problem for realistic situations with a varying noise level. Another disadvantage is the fact that the algorithm cannot adapt to a varying noise level during segments of speech. The adaptive filtering is always based on the estimated noise spectrum of the preceding speech pause.

An improvement of the spectral subtraction technique would be an estimation of the noise spectrum without the necessity of a speech pause detection. A method is presented in this report to estimate the noise spectrum without a speech pause detection.

One application for this method presented in this report is the estimation of the actual SNR of a noisy signal. Furthermore the technique is applied to speech enhancement based on a spectral subtraction respectively on a nonlinear high-pass filtering of the spectral envelopes dependent on the actual SNR.

2. Principal Idea

The principal idea to estimate the noise level in a certain subband is based on a statistical analysis of a segment of the magnitude spectral envelope.

Looking at such a spectral envelope in figure 2.1 and the corresponding distribution density function in figure 2.2 the most commonly occurring spectral magnitude value is zero. The spectral envelope was calculated for a clean speech signal with a duration of about 4.5 s and in a subband of about 500 Hz. The distribution density function was calculated for the whole duration of 4.5 s with an accuracy of about 1 percent in regard to the maximum spectral value inside this subband. The function is shown for the range of 0 to 50 percent of the maximum. Only a few values occur which are higher than 50 percent.

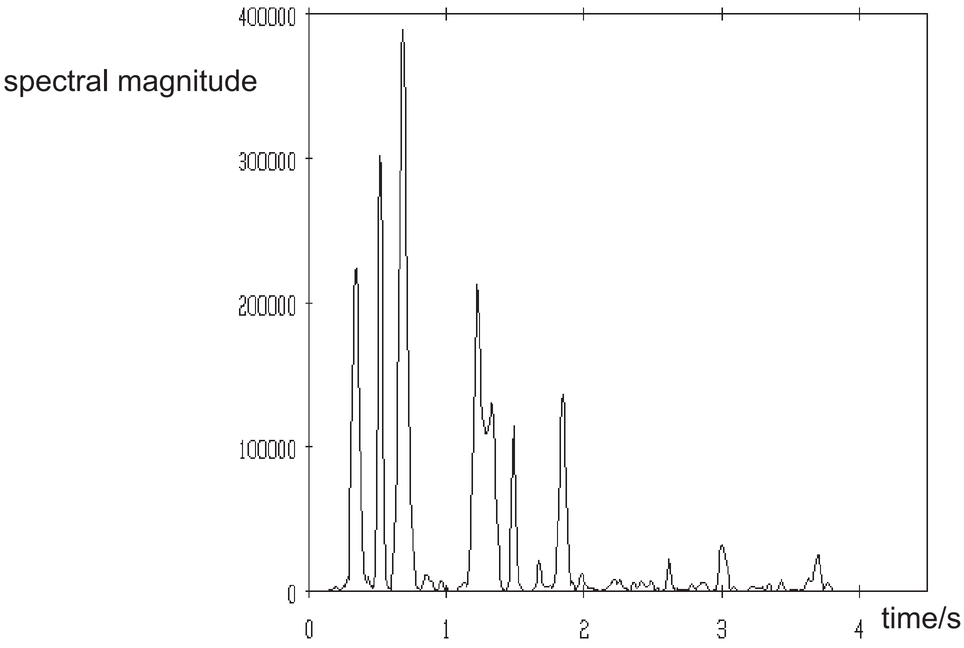

Figure 2.1: Spectral envelope in a band with a centre frequency of 500 Hz

Noise was added artificially to this speech signal to produce different SNRs. The results can be seen in figure 2.3. The noise was a bandpass limited Gaussian noise with a centre frequency of 500 Hz and a bandwidth of 200 Hz.

An increase of the maximum value in the distribution function can be observed for a decreasing SNR. This most frequently occurring value can be taken as an estimation for the noise level inside this band.
Also, an increasing variance of the spectral magnitude values of the noise can be seen for an decreasing SNR. Because of a broad distribution the estimation isn't so accurate

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase®
Smarter legal research.