

Diss. ETH No. 10210

# Contributions to Noise Suppression in Monophonic Speech Signals

A dissertation submitted to the  
SWISS FEDERAL INSTITUTE OF TECHNOLOGY  
ZURICH

for the degree of  
Doktor der technischen Wissenschaften

presented by  
WALTER ETTER  
dipl. El. Ing. ETH  
born 17. November 1958  
citizen of Hundwil (AR)

accepted on the recommendation of  
Prof. Dr. G.S. Moschytz, examiner  
Prof. Dr. J.L. Massey, co-examiner

1993

# Abstract

This thesis is concerned with the problem of removing noise from a noisy speech signal under the following assumptions: The speech is picked up by one microphone; no other signal correlated with the noise or the speech is accessible; the noise and speech are assumed to be uncorrelated; and the noise is assumed to be additive. Various noise types are considered ranging from stationary noise to impulse noise.

Previous studies in suppression of near-stationary noise have mainly been concerned with the noise suppression mechanisms, while little attention was paid to the estimation of the noise parameters. In most studies, the noise parameters were estimated during speech pauses. In this thesis we derive a theoretical limit for an estimation of the noise parameters *during* speech activity. A new algorithm is then presented that is capable of tracking slow changes in short-time stationarity of the noise during speech activity. The algorithm is based on the assumption that each frequency component of the speech signal can be represented by a Markov model. It is shown that this algorithm has adaptation properties superior to previous algorithms.

A new method for noise suppression in the frequency domain, called *noise-adaptive spectral magnitude expansion*, is investigated. This method bridges the gap between companders used to overcome the limited dynamic range of a transmission channel and noise suppression techniques such as spectral subtraction or spectral short-time Wiener filtering. The underlying idea of noise-adaptive spectral magnitude expansion is to adapt the crossover point of the spectral expansion in each frequency channel based on the noise level. When no restrictions are imposed on the expansion characteristic, many types of spectral short-time

filtering, such as spectral subtraction and spectral short-time Wiener filtering, can be considered as special cases of noise-adaptive spectral magnitude expansion.

The thesis also presents the performance limits of noise-suppression techniques that are based on an autoregressive (AR) model of the speech signal, such as the linearized maximum a posteriori (MAP) algorithm. A closed form is derived for this iterative algorithm. It is shown that, in practice, the derived closed-form estimator outperforms the iterative version of the linearized MAP algorithm both in terms of speech quality and computational complexity.

Various aspects of this thesis concern the removal of impulse noise. The standard system for the detection of an unknown noise pulse in a speech signal consists of a whitening filter followed by a threshold detector. The whitening filter can be realized by a short-term and a long-term predictor for the speech signal. However, long-term prediction is often inaccurate due to changes in pitch-frequency causing pitch pulses to be detected as noise pulses. For this reason, a new approach is developed: Instead of a long-term predictor and a detector with a fixed threshold, we use a pitch pulse adaptive threshold detector after the short-term predictor. The threshold of this detector is increased in the immediate vicinity of a pitch pulse. The effectiveness of the new algorithm for the detection of noise pulses caused by scratches on old gramophone records is shown by the reduced false detection and the reduced missed detection.

Once a noise pulse has been detected, the signal at this point can be restored from the neighboring signal by interpolation. We present a new algorithm for the interpolation of a missing signal segment. Unlike past algorithms, our algorithm does not assume the same AR parameters on both sides of the missing segment. The relaxation of this stationarity assumption is essential when the duration of the missing segment in the speech signal is on the order of 10 msec. The algorithm provides the optimal solution to the problem of interpolating a missing segment based on the left-sided and right-sided AR-parameter vector. The solution is optimal in the sense of a minimal residual power of the estimated signal. It is also shown that the performance of this interpolation technique is superior to that of other restoration techniques.

# Kurzfassung

Diese Dissertation behandelt die Unterdrückung von Geräuschen in einem Sprachsignal, welches mit einem Mikrofon aufgenommen wurde. Dabei werden verschiedene Geräuschtypen berücksichtigt, angefangen von stationären Geräuschen bis hin zu Impulsgeräuschen.

Vorangegangene Studien zur Unterdrückung von beinahe-stationären Geräuschen konzentrierten sich in erster Linie auf die Geräuschunterdrückungs-Mechanismen, jedoch selten auf die Schätzung der Geräuschparameter. In den meisten Studien wurden die Geräuschparameter während Sprachpausen geschätzt. In der vorliegenden Arbeit wird zuerst eine theoretische Grenze für die Schätzung der Geräuschparameter während sprachaktiver Phasen hergeleitet. Danach wird ein neuer Algorithmus zur Schätzung der Geräuschparameter präsentiert, welcher in der Lage ist, langsamen Änderungen der Geräuschparameter während sprachaktiver Phasen zu folgen. Der Algorithmus basiert auf der Annahme, dass jede Frequenzkomponente des Sprachsignals mit einem Markov-Modell repräsentiert werden kann. Es wird gezeigt, dass dieser Algorithmus anderen Verfahren bezüglich Adaptionsverhalten überlegen ist.

Unter dem Namen *Noise-Adaptive Spectral Magnitude Expansion* wird eine neue Methode zur Geräuschunterdrückung im Frequenzbereich eingeführt. Sie verbindet die Lücke zwischen Kompandern, die bei begrenzter Dynamik eines Übertragungskanals verwendet werden, und Verfahren zur Geräuschunterdrückung wie *Spektrale Subtraktion* oder *Kurzzeit Wiener-Filterung*. Die zugrundeliegende Idee von *Noise-Adaptive Spectral Magnitude Expansion* ist die dauernde Nachführung des Arbeitspunktes der spektralen Expansion für jeden Frequenzkanal

in Abhängigkeit des Geräuschpegels.

Diese Dissertation präsentiert im weiteren die Leistungsgrenzen von Verfahren zur Geräuschunterdrückung, welche ein autoregressives (AR) Modell der Sprache implizieren. Ein solches Verfahren stellt der linearisierte Maximum A Posteriori (MAP) Algorithmus dar. Für diesen iterativen Algorithmus wird eine geschlossene Form hergeleitet. Es wird gezeigt, dass der auf der geschlossenen Form beruhende Estimator in der Praxis bessere Resultate erzielt als der iterative linearisierte MAP Algorithmus.

Verschiedene Beiträge dieser Arbeit betreffen die Unterdrückung von Impulsgeräuschen. Der Standardansatz zur Detektion von Impulsgeräuschen besteht aus einem Whitening-Filter, das von einem Schwellwert-Detektor gefolgt wird. Das Whitening-Filter kann mit einem Kurzzeit- und einem Langzeit-Prädiktor für das Sprachsignal realisiert werden. Anstelle des Langzeit-Prädictors und eines Detektors mit festem Schwellwert verwenden wir einen Pitch-Pulse-adaptiven Schwellwert-Detektor als Folglied des Kurzzeit-Prädictors. Der Schwellwert wird dabei in unmittelbarer Umgebung eines Pitch-Pulses angehoben. Die Wirkungsweise dieses Algorithmus wird für die Detektion von Geräuschimpulsen demonstriert, welche von Kratzern auf Grammophon-Platten herrühren. Gegenüber dem Standard-Verfahren ergibt sich eine erhebliche Reduktion der Detektionsfehler.

Nachdem ein Impulsgeräusch detektiert wurde, kann das Signal an dieser Stelle durch Interpolation mithilfe von benachbarten Signalanteilen rekonstruiert werden. Es wird ein neues Verfahren zur Interpolation präsentiert, welches im Gegensatz zu früheren Verfahren nicht von den gleichen AR-Parametern auf den beiden Seiten des fehlenden Intervalls ausgeht. Die damit verbundene Lockerung der Stationaritätsannahme ist entscheidend, wenn die Dauer des fehlenden Intervalls im Sprachsignal in der Größenordnung von 10 ms liegt. Der Algorithmus liefert die optimale Lösung des Problems der Interpolation eines fehlenden Segmentes basierend auf dem links- und rechtsseitigen AR-Parameter Vektor. Die Lösung ist optimal im Sinne einer minimalen Residual-Leistung des geschätzten Signals.