





Arbeiten über Digitale Signalverarbeitung

Herausgegeben von Gerhard Schmidt



# Speech Enhancement in Hands-free Systems for Automobile Environments

von

Vasudev Kandade Rajan

2016



Arbeiten über digitale Signalverarbeitung und Systemtheorie

Band 5

**Vasudev Kandade Rajan**

**Speech Enhancement in Hands-free Systems  
for Automobile Environments**

Shaker Verlag  
Aachen 2017

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: Kiel, Univ., Diss., 2017

Copyright Shaker Verlag 2017

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-5362-3

ISSN 2197-7089

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen

Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9

Internet: [www.shaker.de](http://www.shaker.de) • e-mail: [info@shaker.de](mailto:info@shaker.de)

अनेकसंशयोच्छेदति, परोक्षार्थस्य दर्शकम् ।  
सर्वस्य लोचनं शास्त्रं, यस्य नास्त्यन्ध एव सः ॥

It blasts many doubts, foresees what is not obvious |  
Science is the eye of everyone, one who hasn't got it, is like a blind ||

The Rigveda (1500 BCE)





# Acknowledgments

The work presented in the following thesis was performed by me as a doctoral researcher in the Digital Signal Processing and System Theory (DSS) group at Kiel University.

I would like to first express my sincere thanks to my supervisor Prof. Dr.-Ing. Gerhard Schmidt. He has guided me since my master studies, master thesis, and finally in my PhD work. He trusted and provided me with the opportunity to work in his group. I thank him for being available and pointing to the right directions when I needed it the most. He created an open environment in which our individual ideas thrived, at the same time contributing to the support system of the group. He has personally been very kind and has advised me many times in different aspects of my PhD life. My heartfelt gratitude to you for all the support you have provided.

Thanks to paragon AG who have funded a majority of my work presented in this thesis. The feedback provided by the colleagues especially by Prof. Dr.-Ing. Mohamed Krini at paragon, along with the opportunity to visit the campus for important discussions, have significantly contributed to my work.

I would also like to express a special thanks to Prof. Dr.-Ing. Ulrich Heute whom I have followed through his lectures in Advanced Signals and Systems, Advanced Digital Signal Processing, and Speech Signal Processing during my masters. His advice and point of view for different thesis topics have had a long lasting impression on my approach to research.

The entire DSS team of doctoral, master, and bachelor thesis students can never be thanked enough. All the measurements, drives, recordings, pictures, documentation is the understructure on which I was able to perform my work. A special thanks to my group including Dipl.-Ing. Jochen Withopf, Dipl.-Phys. Christian Lücke, M.Sc. Anne Theiß, M. Sc. Christin Baasch, Dipl.-Wirtsch.-Ing. Tim Claussen, Dipl.-Wirtsch.-Ing. Sebastian Rohde, Dipl.-Ing. Torge Rabsch, M. Sc. Jens Reermann, M. Sc. Stephan Senkbeil, M. Sc. Merikan Koyun including colleagues from the former LNS group who have helped me in my work in one way or the other. Many of them who have not just been colleagues but friends have played a vital role in my time at the university.

Finally, last but not the least, I would like to thank my parents, family, and friends around the world whose support cannot be captured in words. They have always believed in my abilities and stood by me at different stages of my life. I would not have reached here without their love and affection.



# Abstract

A new microphone position in the automobile where microphones are placed on the seat belt is available. The pros outweigh the cons of the position which makes it very attractive to be practically used. In order to be able to use this microphone a set of issues are addressed through signal processing methods. Some of the methods presented are improved versions of the existing ones and some methods are new ideas. They are adapted and assessed in the automobile context. The central work of the thesis is a set of speech enhancement algorithms which are applied to the belt microphones. The speech enhancement chapters presented in the thesis form the basic units of a hands-free system.

Belt microphones when integrated into hands-free systems are used to pick up the speech of the passenger/s in the automobile. The microphone signals also contain the echoes of the remote talker which is played back over the loudspeaker in the automobile. This thesis presents an acoustic echo canceller to remove these echoes. The echo canceller must be able not only to remove the echoes to the extent that a transparent conversation is possible, but also to satisfy measures specified by various standards. In order to achieve this an improved way to control the adaptive filters of the echo canceller is presented. The control involves estimation of unmeasurable quantities. Based on theoretically derived optimal quantities an improved step-size control is presented as compared to the existing ones. By utilizing properties of the impulse response such as the inherent delay, slow varying nature, it is shown that the proposed control method outperforms the existing method which is based on the same principle. The chapter also presents a way to deal with the moving of the microphones when the body of the passenger moves. The problem of "room change" is handled through the two coupling factors which are built-in control mechanisms of the step-size control. The step-size control and the room change is evaluated using standard measures under different realistic automobile scenarios.

The speech enhancement of the microphone also deals with the estimation of the background noise in the automobile environment. The noise estimation chapter of the thesis proposes a new noise estimation scheme applicable to the belt microphones. The scheme considers noise scenarios involving nonstationary signals like the sudden change in the noise properties when the window of the automobile is lowered. By tracking long term average, the long term level, and taking into account the short term dynamics of the background noise, a multiplicative constant based scheme is proposed. The basic idea involves the classification of the current state of the noise as either speech, slowly changing noise, or fast changing noise. Based on this classification the long term average estimate is combined with a dynamic weighting mechanism with the instantaneous input spectrum. By doing so it has been shown that it is possible to track the background noise faster and at the same time to avoid tracking speech. This scheme has been compared

---

with two other well known schemes. The evaluation shows that the proposed scheme is the better choice in the evaluated scenarios.

The traditional Wiener filter approach to noise suppression has been re-looked in the final speech enhancement chapter of the thesis. The existing modifications of the Wiener filter are presented as a basis for proposing newer modifications. The first proposed modification moves from the retention of the background noise as a suppressed version to a shaping of the suppressed noise. Two ways to reshape the background noise are proposed, first involving the low frequency noise present in the belt microphones due to its proximity to the windows, second involving the equalization of the noise in the remaining frequencies. The idea is to permanently apply the low frequency modification and apply the broadband equalization to only non-speech frequencies. The second proposed modification tries to reduce the speech distortion caused when the background noise needs to be overestimated to avoid the so-called “musical noise”. The modifications are subjectively tested and the improvements of the methods are shown through spectrogram plots.

A hands-free system is presented where the above proposed speech enhancement algorithms has been implemented in a real-time system. This system has been tested in two cars. The software and hardware implementation details are described in the real-time implementation chapter of the thesis. The evaluation of the hands-free system into which the speech enhancement units are integrated is shown.

# Zusammenfassung

Eine neuartige Mikrofon-Positionierung im Automobil, bei der Mikrofone am Sicherheitsgurt befestigt sind, ist verfügbar. Die Vorteile dieser neuen Positionierung überwiegen deutlich und machen diese gleichzeitig für eine praktische Anwendung sehr attraktiv. Für die Anwendung dieser neuen Positionierung sind eine Reihe von Anforderungen mithilfe von Signalverarbeitungsmethoden zu erfüllen. Einige der vorgestellten Methoden dienen der Verbesserung bestehender Verfahren, wobei auch neue Ideen verfolgt werden. Diese Methoden sind speziell an die Anforderungen im Automobilbereich angepasst. Der Kern dieser Dissertation beschäftigt sich mit einer Reihe von Sprachverbesserungsalgorithmen, welche auf die Gurt-Mikrofone angewendet werden können. Die vorgestellten Algorithmen in den Kapiteln zur Sprachverbesserung bilden die Grundbausteine eines Freisprechsystems.

In Freisprechsystemen werden Gurt-Mikrofone dazu verwendet, die Sprache der Passagiere aufzuzeichnen. Die Mikrofonsignale beinhalten dabei auch die Echos des fernen Sprechers, dessen Signal über die Lautsprecher im Automobil wiedergegeben wird. Um diese Echos aus dem Signal zu entfernen, wird in dieser Arbeit ein akustischer Echo-Kompensator vorgestellt. Der Echo-Kompensator muss dabei in der Lage sein, die Echos so stark zu unterdrücken, dass eine transparente Konversation möglich ist. Zudem sind die Vorgaben der, durch verschiedene Standards spezifizierten, Maße zu erfüllen. Um dies zu erreichen, wird eine verbesserte Steuerung der adaptiven Filter des Echo-Kompensators vorgestellt. Diese schließt dabei die Schätzung von nicht-messbaren Größen ein. Basierend auf theoretisch hergeleiteten, optimalen Größen, wird eine verbesserte Schrittweitensteuerung präsentiert und mit bereits bestehenden Verfahren verglichen. Unter Ausnutzung einiger Eigenschaften der Impulsantwort, wie die dazugehörige Verzögerung und ihre sich nur langsam verändernde Natur, wird gezeigt, dass das vorgestellte Steuerungsverfahren, die bereits existierende, auf demselben Prinzip basierende Methode, übertrifft. Das Kapitel zeigt zusätzlich eine Möglichkeit mit der Bewegung der Mikrofone, auf Grund der Bewegung des Mitfahrers, umzugehen. Das Problem der Raumänderung wird durch die beiden Kopplungsfaktoren, welche eingebaute Steuerungsmechanismen der Schrittweitensteuerung sind, gelöst. Die Schrittweitensteuerung, sowie die Raumänderung werden durch die Verwendung von Standardmaßen unter verschiedenen realistischen Szenarien im Automobil evaluiert.

Zur Sprachverbesserung des Mikrofonsignals gehört zusätzlich die Schätzung des Hintergrundgeräusches in der automobilen Umgebung. Das Kapitel der Geräuschschätzung in dieser Arbeit beschreibt ein neues Verfahren, welches für die Gurt-Mikrofone anwendbar ist. Es berücksichtigt Geräuschszenarien, die instationäre Signale, wie die plötzlichen Änderungen der Geräuscheigenschaften beim Öffnen eines Fensters, einbeziehen. Durch das Tracking des Langzeit-Mittels, des Langzeit-Pegels und unter Berücksichtigung der

---

Kurzzeit-Dynamik des Hintergrundgeräusches, wird ein auf einer multiplikativen Konstante basierendes Verfahren vorgestellt. Vereinfacht ausgedrückt wird bei diesem Verfahren der aktuelle Zustand des Geräusches entweder als Sprache, als sich langsam änderndes Geräusch oder als sich schnell änderndes Geräusch klassifiziert. Basierend auf dieser Klassifikation gibt es eine langzeit Schätzung, die am Ende mit eine dynamischen Gewichtungsfaktor mit dem momentanen Spektrum kombiniert wird. Es konnte gezeigt werden, dass es mit dieser Methode möglich ist, dem Hintergrundgeräusch schnell zu folgen, ohne dabei die Sprache anzugreifen. Diese Geräuschschätzung wurde mit zwei anderen bekannten Verfahren verglichen. Das Ergebnis der Evaluation zeigt, dass, für die ausgewerteten Szenarien, das vorgeschlagene Verfahren die bessere Wahl ist.

Im letzten Kapitel zur Sprachverbesserung wurde noch einmal der traditionelle Wiener-Filter-Ansatz zur Geräuschunterdrückung betrachtet. Bereits bestehende Modifikationen zum Wiener Filter werden, als Basis für neuere Modifikationen, vorgestellt. Bei der ersten vorgeschlagenen Modifikation wird von der Beibehaltung des Hintergrundgeräusches als unterdrücktes Geräusch abgesehen und übergegangen zu einer Modellierung des unterdrückten Geräusches. Dazu werden zwei neue Wege vorgestellt, das Hintergrundgeräusch zu modellieren: Der erste behandelt das Geräusch in den niedrigen Frequenzen, welches bei den Gurt- Mikrofonen durch ihre Nähe zu den Fenstern auftritt, der zweite basiert auf der Entzerrung des Geräusches in den übrigen Frequenzen. Die Idee zur Kombination beider Ansätze ist, die Modifikation in den niedrigen Frequenzen permanent anzuwenden und die breitbandige Entzerrung nur für Frequenzen anzuwenden, die keine Sprache enthalten. Die zweite vorgestellte Modifikation versucht die Störung der Sprache zu vermindern, wenn das Geräusch überschätzt werden muss, um das so genannte "Musical Noise" zu verhindern. Beide Modifikationen wurden subjektiv evaluiert und die Verbesserungen der Methoden durch ein Spektrogramm dargestellt.

Ein Freisprechsystem, welches die vorgestellten Sprachverbesserungsalgorithmen verwendet, wurde in Rahmen einer Echtzeit-Signalverarbeitung implementiert. Das System wurde in zwei verschiedenen Personenkraftwagen getestet. Die Details zur Software- und Hardware-Umsetzung sind im Kapitel der Echtzeit-Umsetzung in dieser Arbeit beschrieben. Schließlich wird die Evaluation des Freisprechsystems, in welches die Sprachverbesserungs-Einheiten integriert sind, gezeigt.

# Contents

<b>Abbreviations and Notation</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background and Processing Framework</b>	<b>5</b>
2.1 Belt Microphones: Microphones Placed on the Seat Belt . . . . .	5
2.1.1 SNR Based Comparison of Belt Microphones . . . . .	6
2.1.2 Challenges Posed by Belt Microphones . . . . .	8
2.2 Processing in the Subband Domain . . . . .	9
2.3 Overview of Processing Components . . . . .	13
<b>3 Multichannel Subband Echo Cancellation</b>	<b>15</b>
3.1 Problem Description . . . . .	15
3.2 The Loudspeaker–Enclosure–Microphone System . . . . .	16
3.2.1 Multichannel LEMS . . . . .	18
3.2.2 Solution to the Problem of Echoes . . . . .	20
3.2.3 The Nonuniqueness Problem in a Multichannel Setup . . . . .	20
3.3 Setup of the Room . . . . .	23
3.3.1 The Echo Path Impulse Response . . . . .	23
3.3.2 Local Speech, Remote Speech, and Background Noise . . . . .	24
3.3.3 Typical Speech Activity Situations . . . . .	25
3.4 Previous Work . . . . .	27
3.5 Echo Path Identification . . . . .	28
3.5.1 System Identification with the NLMS . . . . .	28
3.5.2 Step-Size Control for Echo Cancellation . . . . .	30
3.5.3 Applying the Pseudo-Optimal Step-Size Control . . . . .	34
3.5.4 Activity Detection Methods for Step-Size Control . . . . .	37
3.5.5 Coupling Factors Estimation for Step-Size Control . . . . .	39
3.5.6 Reference Delay Estimation . . . . .	43
3.5.7 NLMS Update with Time-Varying Step-Size . . . . .	44
3.5.8 Nature of the Step-Size Control . . . . .	44
3.6 Residual Echo Estimation . . . . .	46
3.7 Handling System Change Through Coupling Trigger Approach . . . . .	47
3.7.1 Tracking with Coupling Factors . . . . .	48
3.7.2 Trigger Method to Handle Room Changes . . . . .	48
3.7.3 Detection Using Shadow Filter . . . . .	49

3.8	Overview Operation of the Echo Canceller . . . . .	50
3.9	Extending the Single-Channel Setup to a Multichannel Setup . . . . .	51
3.10	Evaluation and Results . . . . .	53
3.10.1	Evaluation of the Control and Detection Methods . . . . .	54
3.10.2	Evaluation Setup . . . . .	54
3.10.3	Performance Measures . . . . .	54
3.10.4	Results . . . . .	56
3.11	Conclusions . . . . .	74
<b>4</b>	<b>A New Low Complexity Noise Estimation Scheme</b>	<b>77</b>
4.1	Introduction . . . . .	77
4.2	Approaches to Reducing Noise . . . . .	82
4.3	Single Channel Noise Power Estimation . . . . .	84
4.3.1	Noisy Signal Model . . . . .	84
4.3.2	Existing Noise Power Estimation Algorithms . . . . .	85
4.4	Design Constraints for the Noise PSD Estimator . . . . .	86
4.5	Simple Noise PSD Estimator . . . . .	87
4.5.1	Basic Principle . . . . .	87
4.5.2	Input Magnitude Spectrum Smoothing . . . . .	87
4.5.3	Noise Tracking . . . . .	88
4.6	Proposed Noise PSD Estimation . . . . .	89
4.6.1	Incrementation Phase . . . . .	90
4.6.2	Decrementation Phase . . . . .	91
4.6.3	Final Multiplicative Constant . . . . .	92
4.6.4	Trend: Long-term Activity Measurement . . . . .	92
4.6.5	Combination of the Pre-estimate and the Input Spectrum . . . . .	93
4.6.6	Summary of the Algorithm . . . . .	95
4.7	Evaluation and Results . . . . .	95
4.7.1	Analyzing the Proposed Algorithm . . . . .	97
4.7.2	Quality Measures . . . . .	105
4.7.3	Results from Real Noise Recordings . . . . .	106
4.8	Conclusion . . . . .	111
<b>5</b>	<b>Improvements in Noise Reduction</b>	<b>115</b>
5.1	Introduction . . . . .	115
5.2	Noise Reduction Filter . . . . .	116
5.3	Wiener Filter as a Noise Reduction Filter . . . . .	117
5.3.1	Tradeoff Between Noise Reduction and Speech Distortion . . . . .	119
5.3.2	Modifications to the Classical Wiener Filter . . . . .	119
5.3.3	Combined Noise and Residual Echo Suppression . . . . .	120
5.4	Residual Noise Shaping . . . . .	120
5.4.1	Shaping the Noise Spectrum . . . . .	121
5.4.2	Desired Noise Shape . . . . .	122
5.4.3	Adjusting the Spectral Floor . . . . .	122



---

5.4.4	Applying the Input-dependent Attenuation . . . . .	124
5.4.5	Residual Noise Shaping Summary . . . . .	126
5.5	Reducing Speech Distortion . . . . .	126
5.5.1	Fixed Overestimation . . . . .	127
5.5.2	Frequency-Dependent Overestimation . . . . .	128
5.5.3	Time-frequency Varying Overestimation . . . . .	129
5.6	Evaluation and Results . . . . .	132
5.6.1	Noise Shaping Through Low-Frequency Compressor . . . . .	132
5.6.2	Noise Shaping Through Broadband Equalizer . . . . .	134
5.6.3	Adaptive Overestimation . . . . .	135
5.7	Conclusion . . . . .	137
<b>6</b>	<b>Real-time Implementation and System Evaluation</b>	<b>141</b>
6.1	Hardware Setup . . . . .	141
6.1.1	Automotive Environment . . . . .	142
6.1.2	Inputs and Outputs . . . . .	142
6.1.3	Processing Unit . . . . .	143
6.2	Solution Architecture . . . . .	144
6.2.1	Processing Components . . . . .	145
6.3	Software Realization through KiRAT . . . . .	149
6.4	Evaluation of the Hands-free System on a Testbench . . . . .	150
6.4.1	Double-Talk Tests . . . . .	153
6.5	Conclusion . . . . .	154
<b>7</b>	<b>Summary and Future Work</b>	<b>155</b>
7.1	Future Work . . . . .	157
	<b>List of Figures</b>	<b>e</b>
	<b>List of Tables</b>	<b>g</b>
	<b>A Noise Recordings for the Belt Microphones</b>	<b>i</b>
	<b>Bibliography</b>	<b>o</b>