

A microscopic model of speech recognition for listeners with normal and impaired hearing

Von der Fakultät für Mathematik und Naturwissenschaften
der Carl-von-Ossietzky-Universität Oldenburg
zur Erlangung des Grades und Titels eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

Dipl.-Phys. Tim Jürgens
geboren am 25. Mai 1979
in Wilhelmshaven

Erstgutachter: Prof. Dr. Dr. Birger Kollmeier

Zweitgutachter: PD Dr. Volker Hohmann

Tag der Disputation: 25. November 2010

Berichte aus der Medizinischen Physik

Tim Jürgens

**A microscopic model of
speech recognition for listeners with
normal and impaired hearing**

Shaker Verlag
Aachen 2011

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.: Oldenburg, Univ., Diss., 2010

Copyright Shaker Verlag 2011

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 978-3-8440-0147-1

ISSN 1617-2965

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

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

Internet: www.shaker.de • e-mail: info@shaker.de

für Andreas

Abstract

Degraded speech intelligibility is one of the most frequent complaints of sensorineural hearing-impaired listeners, both in noisy and quiet situations. An understanding of the effect of hearing impairment on speech intelligibility is therefore of large interest particularly in order to develop new hearing-aid algorithms for rehabilitation. However, sensorineural hearing impairment is often found to be very individual in terms of the functional deficits of the inner ear and the entire auditory system. Important individual factors to be considered when modeling the effect of sensorineural hearing impairment on speech intelligibility are the audibility of the speech signal, different compressive properties, or different active processes in the inner ear. The latter two can be termed *supra-threshold* factors, since they affect the processing of speech well above the individual absolute threshold. It is not possible to directly (i.e. invasively) measure and study the influence of these supra-threshold factors on human speech recognition (HSR) for ethical reasons. However, computer models on HSR can provide an insight in how these factors may influence speech recognition performance.

This dissertation presents a microscopic model of human speech recognition, microscopic in a sense that first, the recognition of single phonemes rather than the recognition of whole sentences is modeled. Second, the particular *spectro-temporal* structure of speech is processed in a way that is presumably very similar to the processing that takes place in the human auditory system. This contrasts with other models of HSR, which usually use the *spectral* structure only. This microscopic model is capable of predicting phoneme recognition in normal-hearing listeners in noise (Chapter 2) along with important aspects of consonant recognition in normal-hearing and hearing-impaired listeners in quiet condition (Chapter 5). Furthermore, an extension of this model for the prediction of word recognition rates in whole German sentences is capable of predicting speech reception thresholds of normal-hearing and hearing-impaired listeners as accurately as a standard speech intelligibility model (Chapter 3). Parameters reflecting the supra-threshold auditory processing are assessed in normal-hearing and hearing-impaired listeners using indirect psychoacoustical measurement techniques such as a forward masking experiment and categorical loudness scaling (Chapter 4). Finally, the influence of including supra-threshold auditory processing deficits (assessed using the aforementioned measurement techniques) in modeling speech recognition is investigated (Chapter 5) primarily realized as a loss in cochlear compression. The results show that implementing supra-threshold processing deficits (as found in hearing-impaired listeners) in a

microscopic model of human speech recognition improves prediction accuracy. However, the advantage of taking these additional suprathreshold processing parameters into account is marginal in comparison to predicting speech intelligibility directly from audiometric data.

Zusammenfassung

Eins der Hauptprobleme von Leuten mit einer Schallempfindungsschwerhörigkeit ist eine verschlechterte Sprachverständlichkeit sowohl in Ruhe, als auch in Umgebungen mit Störgeräusch. Ein Verständnis davon zu gewinnen, wie Schwerhörigkeit Sprachverständlichkeit beeinflusst, ist daher von großer Wichtigkeit für die Rehabilitation Schwerhörender, z.B. in Form der Entwicklung neuer Hörgeräte-algorithmen. Schallempfindungsschwerhörigkeit kann allerdings sehr individuell sein, wenn man die Art und Anzahl der geschädigten Komponenten des Innenohres und des gesamten auditorischen Systems betrachtet. Wichtige individuelle Faktoren der Schallempfindungsschwerhörigkeit, welche Sprachverständlichkeit beeinflussen, können zum Beispiel sein: die Hörbarkeit des Sprachsignals, unterschiedliche kompressive Eigenschaften in der Verarbeitung des Innenohres oder unterschiedlich starke aktive Prozesse im Innenohr. Die letzteren beiden können als *überschwellige* Faktoren bezeichnet werden, da sie die Verarbeitung von Sprache oberhalb der Hörschwelle beeinflussen. Es ist aus ethischen Gründen nicht möglich den Einfluss dieser überschwelligen Faktoren auf die menschliche Spracherkennung direkt (also invasiv) zu messen und zu studieren. Allerdings können Computermodelle der menschlichen Spracherkennung einen Einblick geben, wie diese Faktoren die Sprachverständlichkeitsleistung beeinflussen können.

Diese Dissertation präsentiert ein mikroskopisches Modell der menschlichen Spracherkennung, mikroskopisch in dem Sinne, dass erstens die Erkennung von einzelnen Phonemen anstelle der Erkennung von ganzen Wörtern oder Sätzen modelliert wird. Zweitens wird die genaue *spekro-temporale* Struktur von Sprache auf eine Art und Weise verarbeitet, die sehr ähnlich zu der Verarbeitung ist, wie sie auch im menschlichen auditorischen System stattfindet. Andere gängige Modelle der menschlichen Spracherkennung nutzen im Gegensatz dazu nur die *spektrale* Struktur von Sprache und einem optionalen Störgeräusch aus. Dieses mikroskopische Modell ist dazu in der Lage Phonemerkennungsraten für Normalhörende unter Einfluss von Hintergrundrauschen (Kapitel 2) und wichtige Aspekte der Konsonanterkennung für Normal- und Schwerhörende in Ruhe (Kapitel 5) vorherzusagen. Außerdem kann eine Erweiterung dieses Modells auf die Erkennung von Wörtern (eingebettet in ganzen deutschen Sätzen) die Sprachverständlichkeitsschwellen von Normal- und Schwerhörenden mit ebenso großer Genauigkeit vorhersagen wie ein anderes gängiges Sprachverständlichkeitsmodell (Kapitel 3). Parameter, die die überschwellige auditorische Verarbeitung in Normal- und Schwerhörenden quantifizieren, wurden mit Hilfe von indirekten psychoakustischen

Messungen, nämlich einem Nachverdeckungsexperiment und der kategorialen Lautheitsskalierung geschätzt (Kapitel 4). In Kapitel 5 wurde dann schlussendlich untersucht, welchen Einfluss eine Veränderung der überschwelligen Verarbeitung (geschätzt aus den Messungen aus Kapitel 4) auf die modellierte Sprachverständlichkeit hat. Die Ergebnisse zeigen, dass der Einbau einer überschwelligen Verarbeitung, so wie sie in Schwerhörenden beobachtet wird, die Vorhersage der Sprachverständlichkeit verbessert. Allerdings ist der Vorteil, der durch den Einbau der genauen überschwelligen Verarbeitung (geschätzt durch überschwellige psychoakustische Messungen) erreicht wird, marginal im Gegensatz zu einer alleinigen Schätzung dieser überschwelligen Verarbeitung durch das Audiogramm.

List of publications associated with this thesis

Peer-reviewed articles:

Jürgens, T., Brand, T., Kollmeier, B. (2007), “Modelling the human-machine gap in speech reception: microscopic speech intelligibility prediction for normal-hearing subjects with an auditory model,” Proceedings of the 8th annual conference of the International Speech Communication Association (Interspeech, Antwerp, Belgium), pp. 410-413.

Jürgens, T., Brand, T. (2009), “Microscopic prediction of speech recognition for listeners with normal hearing in noise using an auditory model,” J. Acoust. Soc. Am. 126, pp. 2635-2648.

Jürgens, T., Fredelake, S., Meyer, R. M., Kollmeier, B., Brand, T. (2010), “Challenging the Speech Intelligibility Index: macroscopic vs. microscopic prediction of sentence recognition in normal and hearing-impaired listeners,” Proceedings of the 11th annual conference of the International Speech Communication Association (Interspeech, Makuhari, Japan), pp. 2478-2481.

Jürgens, T., Kollmeier, B., Brand, T., Ewert, S.D. (2010), “Assessment of auditory nonlinearity for listeners with different hearing losses using temporal masking and categorical loudness scaling,” submitted to Hear. Res.

Non-peer-reviewed articles:

Jürgens, T., Brand, T., Kollmeier, B. (2007), “Modellierung der Sprachverständlichkeit mit einem auditorischen Perzeptionsmodell,” Tagungsband der 33. Jahrestagung für Akustik (DAGA, Stuttgart, Germany), pp. 717-718.

Jürgens, T., Brand, T., Kollmeier, B. (2008), “Sprachverständlichkeitsvorhersage für Normalhörende mit einem auditorischen Modell,” Tagungsband der 11. Jahrestagung der Deutschen Gesellschaft für Audiologie (DGA, Kiel, Germany).

Jürgens, T., Brand, T., Kollmeier, B. (2008), “Phonemerkennung in Ruhe und im Störgeräusch, Vergleich von Messung und Modellierung,” Tagungsband der 39. Jahrestagung der Deutschen Gesellschaft für Medizinische Physik (DGMP, Oldenburg, Germany).

Jürgens, T., Brand, T., Kollmeier, B. (2009), “Consonant recognition of listeners with hearing impairment and comparison to predictions using an auditory model,” Proceedings of the NAG/DAGA International Conference on Acoustics (Rotterdam, The Netherlands), pp. 1663-1666.

Jürgens, T., Brand, T., Ewert, S. D., Kollmeier, B. (2010), “Schätzung der Nichtlinearität der auditorischen Verarbeitung bei Normal- und Schwerhörenden durch kategoriale Lautheitsskalierung,” Tagungsband der 36. Jahrestagung für Akustik (DAGA, Berlin, Germany), pp. 467-468.

Published abstracts:

Jürgens, T., Brand, T., Kollmeier, B. (2009), "Predicting consonant recognition in quiet for listeners with normal hearing and hearing impairment using an auditory model," J. Acoust. Soc. Am. **125**, p. 2533 (157th meeting of the Acoustical Society of America, Portland, Oregon).

Contents

Abstract	5
Zusammenfassung	7
List of publications associated with this thesis.....	9
Contents.....	11
1 General Introduction	17
2 Microscopic prediction of speech recognition for listeners with normal hearing in noise using an auditory model.....	23
2.1 Introduction	24
2.1.1 Microscopic modeling of speech recognition	25
2.1.2 A-priori knowledge	27
2.1.3 Measures for perceptual distances	27
2.2 Method	28
2.2.1 Model structure	28
2.2.2 Speech corpus	33
2.2.3 Test conditions	34
2.2.4 Modeling of a-priori knowledge	34
2.2.5 Subjects	35
2.2.6 Speech tests	35
2.3 Results and discussion	36
2.3.1 Average recognition rates	36
2.3.2 Phoneme recognition rates at different SNRs	38
2.3.3 Phoneme confusion matrices	39
2.4 General discussion	44
2.4.1 Microscopic prediction of speech intelligibility	44
2.4.2 Distance measures	46
2.4.3 Phoneme recognition rates and confusions	47
2.4.4 Variability in the data	49
2.4.5 Practical relevance	49

2.5	Conclusions	50
2.6	Acknowledgements	50
2.7	Appendix: Significance of confusion matrices elements	51
3	Challenging the Speech Intelligibility Index: Macroscopic vs. microscopic prediction of sentence recognition in normal and hearing-impaired listeners.	53
3.1	Introduction	54
3.2	Measurements	54
3.2.1	Subjects	54
3.2.2	Apparatus	55
3.2.3	Speech intelligibility measurements	55
3.3	Modeling	56
3.3.1	Speech Intelligibility Index	56
3.3.2	Microscopic model	57
3.4	Results and comparison	59
3.5	Discussion	61
3.6	Conclusions	62
3.7	Acknowledgements	63
4	Assessment of auditory nonlinearity for listeners with different hearing losses using temporal masking and categorical loudness scaling	65
4.1	Introduction	66
4.2	Method	69
4.2.1	Subjects	69
4.2.2	Apparatus and calibration	70
4.2.3	Procedure and stimuli	70
4.3	Experimental results	74
4.3.1	Temporal masking curves	74
4.3.2	Categorical loudness scaling data	76
4.4	Data analysis and comparison	77
4.4.1	Estimates of low-level gain, gain loss, and compression ratio from TMC	77
4.4.2	Estimates of inner and outer hair cell loss from off-frequency TMCs	79
4.4.3	Estimates of HL_{OHC} from ACALOS	81
4.4.4	Comparison of parameters derived from TMCs and ACALOS	85

4.4.5 Variability of parameters	87
4.5 Discussion	89
4.5.1 Possible systematic deviations of parameters derived from TMCs	89
4.5.2 Relation of ACALOS loudness functions to classical loudness functions	91
4.5.3 Possible systematic deviations of parameters from ACALOS	92
4.5.4 Correlation of parameters derived from TMCs and ACALOS	93
4.6 Conclusions	95
4.7 Acknowledgements	96
4.8 Appendix: Data of a listener with combined conductive and sensorineural hearing loss	97
5 Prediction of consonant recognition in quiet for listener with normal and impaired hearing using an auditory model	99
5.1 Introduction	100
5.2 Experiment I: phoneme recognition in normal-hearing listeners	103
5.2.1 Method	103
5.3 Experiment II: consonant recognition in hearing-impaired listeners	106
5.3.1 Method	106
5.4 Estimation of individual supra-threshold processing	107
5.5 Modeling human speech recognition	109
5.5.1 Microscopic speech recognition model	109
5.5.2 Model versions to implement hearing impairment	109
5.6 Comparison of observed and predicted results	115
5.6.1 Modeling data of Experiment I	116
5.6.2 Modeling data of Experiment II	119
5.7 General discussion	126
5.7.1 Audibility	127
5.7.2 Compression	128
5.7.3 Phoneme recognition rates and confusions	131
5.7.4 Sensorineural hearing impairment	136
5.8 Conclusions	138
5.9 Acknowledgements	139
5.10 Appendix	140
5.10.1 Vowel recognition of normal-hearing listeners	140

5.10.2	Relations between speech recognition and compression in hearing-aid studies.....	142
6	Summary and concluding remarks.....	145
7	Appendix: Modeling the human-machine gap in speech reception: microscopic speech intelligibility prediction for normal-hearing subjects with an auditory model.....	151
7.1	Introduction.....	152
7.2	Measurements	152
7.2.1	Method	152
7.2.2	Results	153
7.3	The perception model.....	155
7.3.1	Specification.....	155
7.3.2	Model predictions and comparison with listening tests	157
7.4	Discussion	159
7.5	Conclusions	160
7.6	Acknowledgements	161
8	Bibliography.....	163
9	Danksagung.....	173
10	Lebenslauf	177
11	Erklärung	178
12	List of abbreviations.....	179