

Nonnegative Matrix Factorization with Adaptive Elements for Monaural Audio Source Separation

**Von der Fakultät für Elektrotechnik und Informationstechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors der
Ingenieurwissenschaften genehmigte Dissertation**

vorgelegt von
Diplom-Ingenieur
Julian Mathias Becker
aus Bad Kreuznach

Berichter:
Univ.-Prof. Dr.-Ing. Jens-Rainer Ohm
Univ.-Prof. Dr. rer. nat. Michael Vorländer

Tag der mündlichen Prüfung: 04.08.2016

Aachen Series on Multimedia and Communications Engineering

Volume 16

Julian Becker

**Nonnegative Matrix Factorization with Adaptive
Elements for Monaural Audio Source Separation**

Shaker Verlag
Aachen 2016

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.: D 82 (Diss. RWTH Aachen University, 2016)

Copyright Shaker Verlag 2016

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 978-3-8440-4814-8

ISSN 1614-7782

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

Vorwort

Die vorliegende Dissertation entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Nachrichtentechnik der Rheinisch-Westfälischen Technischen Hochschule Aachen. Ich möchte mich an dieser Stelle bei allen bedanken, durch deren Hilfe und Unterstützung mir der Weg zur Fertigstellung dieser Arbeit erleichtert wurde:

Zunächst möchte ich mich bei Herrn Prof. Dr.-Ing. Jens-Rainer Ohm, dem Leiter des Instituts für Nachrichtentechnik, für die Betreuung meiner Arbeit und die dabei stets positive Arbeitsatmosphäre bedanken. Herrn Prof. Dr. rer. nat. Michael Vorländer danke ich für die Übernahme des Zweitgutachtens und das Interesse an meiner Arbeit.

Es gibt viele weitere Personen, die auf verschiedenste Arten zum Gelingen dieser Dissertation beigetragen haben. Erwähnenswert ist mit Sicherheit die außergewöhnlich gute, kollegiale Stimmung am Institut, die mir sehr geholfen hat, immer motiviert meiner Arbeit nachzugehen. Dafür und auch für den hilfreichen wissenschaftlichen Austausch möchte ich mich bei allen Kollegen am Institut bedanken. Ich bin mir sicher, dass der Kontakt zu vielen Kollegen privat weiter erhalten bleibt.

Ich möchte mich auch bei denjenigen außerhalb des Instituts bedanken, die es mir durch ihre Unterstützung ermöglicht haben, mich an diese Aufgabe heranzuwagen. Vielen Dank auch an alle, deren Ratschläge oder Korrekturen zur Fertigstellung dieser Arbeit beigetragen haben.

Mein persönlicher Werdegang wurde überhaupt erst durch die Unterstützung meiner Eltern ermöglicht. Bei all meinen Entscheidungen hatte ich immer das Gefühl, ihr vollstes Vertrauen zu genießen und konnte auf diese Art selbstständig meinen eigenen Weg gehen. Vielen Dank dafür.

Zu guter Letzt möchte ich mich bei Sophia dafür bedanken, dass sie es mir durch ihre liebevolle Unterstützung ermöglicht hat, auch in arbeitsintensiveren Phasen ruhig und entspannt zu bleiben. Glücklicherweise reihte sich die Geburt unseres Sohnes Janosch sehr passend in meine Zeitplanung ein. Deshalb gilt mein abschließender Dank Janosch für diese zusätzliche Motivation, die Arbeit pünktlich fertig zu stellen.

Aachen, im August 2016

Contents

1	Introduction	1
1.1	Main Contributions of this Thesis	2
1.2	Outline	3
2	Fundamentals	5
2.1	Audio Signal Representations	5
2.1.1	Digital Audio Signals in Time Domain	5
2.1.2	Fourier Transform	6
2.1.3	Time-Frequency Representation	7
2.2	Mel Scale	11
2.3	Harmonic and Percussive Audio Signals	12
2.3.1	Harmonic Ratio	13
2.4	Singular Value Decomposition	14
2.5	Nonnegative Matrix Factorization	15
2.5.1	NMF Model	15
2.5.2	Cost Function	16
2.5.3	Update Rules	16
2.6	State-of-the-Art of Blind Audio Source Separation	17
2.6.1	Time-Frequency Masking	18
2.6.2	Algorithms Using TF-Masking	18
2.6.3	NMF for Source Separation	19
2.6.4	Oracle Estimators	21
2.6.5	Single-channel Source Separation Beyond Time-Frequency Masking	21
2.7	Quality Measures for Source Separation	22
2.7.1	Energy Ratio Measures	23
2.7.2	Perceptually Motivated Measures	24
2.8	Audio Data	25
3	Reference Algorithm - Preliminary Experiments	27
3.1	Unsupervised Source Separation Algorithm	27
3.2	Influence of Parameters	31
3.2.1	Influence of the Time-Frequency Transform	32
3.2.2	Influence of the Parameters of NMF Preprocessing	34
3.2.3	Influence of Nonnegative Matrix Factorization (NMF) Parameters	35
3.3	Summary	38
4	State-Of-The-Art of NMF Extensions for Unsupervised Monaural Audio Source Separation	41
4.1	Limitations of NMF	41
4.1.1	Acoustical Events with Temporal Structure	42

4.1.2	Temporal and Spectral Properties of Audio Signals	43
4.2	Nonnegative Matrix Factor Deconvolution	44
4.2.1	Mathematical Description	44
4.2.2	Application of Nonnegative Matrix Factor Deconvolution (NMFD) to Source Separation	45
4.2.3	Experiments	46
4.2.4	Summary	47
4.3	NMF with Additional Constraints	48
4.3.1	Mathematical Description	48
4.3.2	Squared Difference Smoothness Constraint	49
4.3.3	Experiments	52
4.3.4	Summary	56
4.4	Summary	56
5	Adaptive constraints for NMF	57
5.1	Analysis of the Effect of Constraints	57
5.1.1	Effect of weights	57
5.1.2	Effect of constraints on NMF components	61
5.2	Component-Adaptive Constraints	62
5.2.1	Weighting with the Cost Function	63
5.2.2	Generalization of the Component-Adaptive Constraints	66
5.2.3	Experimental Results	69
5.3	Weight Adaptation	78
5.3.1	Adaptation Algorithm	80
5.3.2	Experimental Results	81
5.4	Summary	86
6	Generalization of the Nonnegative Matrix Factor Deconvolution	89
6.1	Analysis of Problems of the NMFD	89
6.2	Generalized Nonnegative Matrix Factor Deconvolution	93
6.2.1	Experimental Results	96
6.3	Exemplary Algorithm Using the Generalized NMFD	100
6.3.1	Motivation of the algorithm	100
6.3.2	Algorithm description	102
6.3.3	Experimental Results	104
6.4	Summary	110
7	Summary and Conclusions	113
7.1	Summary	113
7.2	Outlook	114
A	Detailed Description of Audio Data	117
A.1	Test Set \mathcal{A}	117
A.2	Test Set \mathcal{B}	119

B Results of the Perceptual Evaluation methods for Audio Source Separation (PEASS) measures	123
B.1 PEASS results to Chapter 4	123
B.1.1 PEASS results to Section 4.2	123
B.1.2 PEASS results to Section 4.3	124
B.2 PEASS results to Chapter 5	125
B.2.1 PEASS results to Section 5.2	125
B.2.2 PEASS results to Section 5.3	128
B.3 PEASS results to Chapter 6	129
B.3.1 PEASS results to Section 6.2.1	129
B.3.2 PEASS results to Section 6.3.3	131
Bibliography	133