

Cambridge University Press
0521619386 - Modelling Auditory Processing and Organisation
Martin Cooke
Frontmatter
[More information](#)

MODELLING AUDITORY PROCESSING AND ORGANISATION

Cambridge University Press
0521619386 - Modelling Auditory Processing and Organisation
Martin Cooke
Frontmatter
[More information](#)

Distinguished Dissertations in Computer Science

Edited by
C.J. van Rijsbergen, University of Glasgow

The Conference of Professors of Computer Science (CPCS) in conjunction with the British Computer Society (BCS), selects annually for publication up to four of the best British Ph.D. dissertations in computer science. The scheme began in 1990. Its aim is to make more visible the significant contribution made by Britain - in particular by students - to computer science, and to provide a model for future students. Dissertations are selected on behalf of CPCS by a panel whose members are:

M. Clint, Queen's University, Belfast
R.J.M. Hughes, University of Glasgow
R. Milner, University of Edinburgh (Chairman)
K. Moody, University of Cambridge
M.S. Paterson, University of Warwick
S. Shrivastava, University of Newcastle upon Tyne
A. Sloman, University of Birmingham
F. Sumner, University of Manchester

Cambridge University Press
0521619386 - Modelling Auditory Processing and Organisation
Martin Cooke
Frontmatter
[More information](#)

MODELLING AUDITORY PROCESSING AND ORGANISATION

Martin Cooke
University of Sheffield



Cambridge University Press
0521619386 - Modelling Auditory Processing and Organisation
Martin Cooke
Frontmatter
[More information](#)

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Cambridge University Press 1993

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1993
First paperback edition 2005

A catalogue record for this book is available from the British Library

ISBN 0 521 45094 2 hardback
ISBN 0 521 61938 6 paperback

Cambridge University Press
0521619386 - Modelling Auditory Processing and Organisation
Martin Cooke
Frontmatter
[More information](#)

for Jo, Sam and Anna

Contents

Acknowledgments	xi
1 Introduction	1
1.1 Segregation and fusion in hearing	1
1.2 Signals and symbols in machine perception	1
1.3 Theoretical influences	3
Marr	3
Marrian influences in speech and hearing	3
1.4 Auditory scene analysis	4
1.5 Previous approaches to sound source segregation	5
Double vowel separation	6
Weintraub's system	6
Discussion	7
1.6 A guide to the thesis	8
2 The auditory periphery: physiology, function and a computer model	9
2.1 Introduction	9
2.2 The structure and function of the auditory periphery	9
Overview	9
Mechanical and tuning properties	10
Hair cell transduction	11
Auditory nerve fibre responses	11
2.3 Computer models of the auditory periphery	13
2.4 A model of the auditory periphery	14
Outline	14
The gammatone filterbank	16
A new approach to gammatone filterbank implementation	17
Comparisons	18

	Analysis of results	21	
	Hair cell static nonlinearity	22	
	Hair cell model	23	
2.5	The central auditory system	30	
	Auditory anatomy	30	
	Responses, maps and function	31	
3	Auditory representations		33
3.1	Introduction	33	
	The composition of the auditory scene	33	
3.2	Computation of synchrony strands	35	
	Overview	35	
	Estimation of dominant components	35	
	Instantaneous frequency	36	
	Smoothed instantaneous frequency	38	
	Calculation of place-groups	38	
	Temporal aggregation of place-groups	42	
	Estimation of strand parameters	42	
	Role of each free parameter in the model	43	
3.3	Validation of the representation by resynthesis	44	
	Motivation	44	
	Resynthesis procedure	44	
	Informal listening results	45	
3.4	A variety of synchrony strand displays	45	
3.5	Discussion	48	
	Limitations and possible improvements	48	
4	Modelling auditory scene exploration		51
4.1	Introduction	51	
4.2	The computational grouping problem faced by the auditory system	51	
	Computational issues	52	
	Search issues: optimal or just 'good' solutions?	54	
	A framework for auditory scene exploration	55	
4.3	Stage 1: The independent application of grouping principles	55	
	Heuristics for search	56	
	Constraint propagation	56	
4.4	Stage 2: Combining groups	59	
	A taxonomy for interpreting pairs of groups	60	
	Combining groups in the model	61	
4.5	Discussion	62	
	Limitations of the model	62	
5	Implementation of auditory grouping principles		63
5.1	Introduction	63	
5.2	Harmonicity	63	
	Experimental findings	63	

Implementation	64	
Discussion	67	
5.3 Common amplitude modulation	68	
Experimental findings	68	
Implementation	68	
5.4 Other principles for simultaneous grouping	70	
Common frequency modulation and movement	70	
Onset and offset synchrony	72	
5.5 Higher-level grouping	73	
Derived properties	73	
Group combination	73	
5.6 Grouping of natural speech signals	75	
5.7 Discussion	75	
6 An evaluation of sound source separation in the model		79
6.1 Introduction	79	
6.2 Evaluation methodologies	79	
6.3 The mixture database	80	
6.4 The mixture correspondence problem	83	
6.5 Metrics	86	
6.6 Procedure	87	
Grouping at random	88	
Grouping with no intrusions	88	
6.7 Results	88	
Utterance characterisation	88	
Positive evidence for the voiced source	89	
Crossover from intrusive source	89	
6.8 Illustrations of grouping	90	
6.9 Discussion	90	
7 Conclusions and future development		93
7.1 Summary of system	93	
7.2 Novelty of the approach	93	
7.3 Limitations of the model	95	
7.4 Parsimony of explanation: The default condition of organisation	95	
7.5 Further development of the model	96	
Appendices		
A Filter derivations		97
A.1 Pole-mapping	97	
A.2 Bilinear transform	98	
A.3 Impulse invariant transform	99	
A.4 Ideal responses	100	

A.5 Digital filter characteristics	100	
B Derivations relating to the hair cell model		103
B.1 Depletion and recovery for constant input	103	
B.2 Time constants of adaptation and recovery	104	
C Derivation of instantaneous frequency		105
Bibliography		109
Index		119

Acknowledgments

It is a pleasure to thank all those individuals whose research has been influential in shaping this work. The strong auditory community of the UK is sufficiently cohesive to provide great help to anyone setting out on a modelling study such as the one described here. I have enjoyed productive discussions with research groups at the Universities of Cambridge (Roy Patterson and John Holdsworth), Edinburgh (Mark Terry), Keele (Bill Ainsworth and Ted Evans), Kiel (Michel Scheffers), Loughborough (Ray Meddis and Michael Hewitt), Nottingham (Quentin Summerfield and Richard Stubbs), Southampton (Bob Dampier) and Sussex (Chris Darwin, Valter Ciocca and Roy Gardner), and with Roger Moore of the Speech Research Unit, RSRE.

I found collaboration with members of the Laboratoire D'Informatique pour la Mecanique et les Sciences de l'Ingenieur in Paris (Christophe D'Alessandro, Jean-Sylvain Lienard and Maxine Eskenazi) in the 1989 Cochleagram-Reading Workshop stimulating; in particular, the use of resynthesis was prompted by the work of Jean-Sylvain Lienard.

My interest in auditory modelling was inspired by David Schofield, a colleague of earlier days at the National Physical Laboratory, whose work in matching physiological and psycho-physical accounts of frequency selectivity has, I believe, proved catalytic for later developments of the gammatone filterbank.

Thanks to: Steve Beet and Robin Sharpe for providing a good deal more insight into filter design and signal processing than I could have obtained from textbooks alone; Chris Darwin for the /ru/-/li/ stimulus used to illustrate grouping in Chapter 5; to the Centre for Speech Technology Research at Edinburgh University and the Department of Linguistics and Phonetics at Leeds for some of the utterances used; and to Lori Lamel for the loan of TIMIT.

Locally, I have enjoyed working with members of SPLASH, the SPeech LABoratory at SHeffield; particular credit must go to Malcolm 'disc-crasher' Crawford, who, apart from many stimulating discussions and all things confectionery, forced me to hit the return key on many crucial processes, when I preferred to pause contemplatively for a few minutes before discovering that an idea was not going to work. Thanks also to Guy Brown for proof-reading beyond the call and useful suggestions. Other members of the Department of Computer Science have, with a mixture of ribaldry and good-humour, made Sheffield a great place to work.

Special thanks are due to my supervisor, Phil Green, who had this work in mind from as long ago as 1984. His encouragement and confidence helped tremendously throughout.

Cambridge University Press
0521619386 - Modelling Auditory Processing and Organisation
Martin Cooke
Frontmatter
[More information](#)

xii

Acknowledgments

The material presented in chapter 3 is re-expressed from a paper published in the journal *Computer Speech and Language* and is used with permission from Academic Press Ltd.