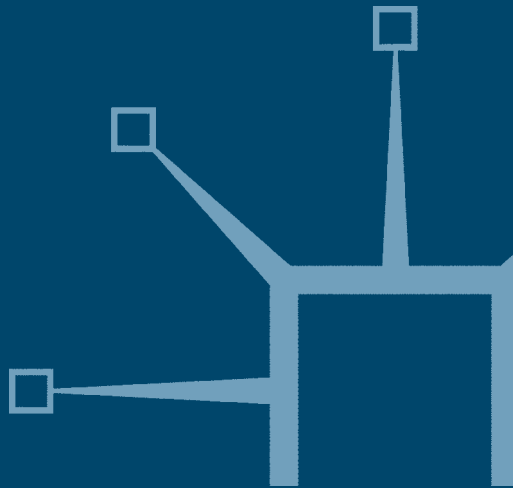


palgrave
macmillan

Speech Production and Perception

Mark Tatham and Katherine Morton



Speech Production and Perception

Other books by Mark Tatham and Katherine Morton

EXPRESSION IN SPEECH: Analysis and Synthesis

DEVELOPMENTS IN SPEECH SYNTHESIS

Speech Production and Perception

Mark Tatham
and
Katherine Morton

palgrave
macmillan



© Mark Tatham and Katherine Morton 2006

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No paragraph of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London W1T 4LP.

Any person who does any unauthorised act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The authors have asserted their rights to be identified as the authors of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published in 2006 by
PALGRAVE MACMILLAN
Houndmills, Basingstoke, Hampshire RG21 6XS and
175 Fifth Avenue, New York, N.Y. 10010
Companies and representatives throughout the world.

PALGRAVE MACMILLAN is the global academic imprint of the Palgrave Macmillan division of St. Martin's Press, LLC and of Palgrave Macmillan Ltd. Macmillan® is a registered trademark in the United States, United Kingdom and other countries. Palgrave is a registered trademark in the European Union and other countries.

ISBN 978-1-4039-1733-1 ISBN 978-0-230-51396-9 (eBook)

DOI 10.1057/9780230513969

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources.

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

Library of Congress Cataloging-in-Publication Data

Tatham, Mark.

Speech production and perception / Mark Tatham and Katherine Morton.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-4039-1732-9 (cloth) – ISBN 978-1-4039-1733-7 (pbk.)

1. Speech. 2. Phonetics. 3. Speech perception. I. Morton, Katherine.

II. Title.

P95.T32 2006

414'.8—dc22

2005044661

10 9 8 7 6 5 4 3 2 1
15 14 13 12 11 10 09 08 07 06

Dedicated to the memory of Peter Ladefoged, 1925–2006

This page intentionally left blank

Contents

List of Figures and Tables xiv

Introduction xvi

Part I Speech Production Theory

1 Classical Phonetics	3
Introduction	3
Assumptions of Classical Phonetics	3
The claims of Classical Phonetics	6
System in Classical Phonetics	7
Parameters for classifying segments	9
Combinatorial constraints on phonetic features	10
Segments ‘in’ the speech signal	11
What is a segment? – the speaker focus	13
What is a segment? – the listener focus	14
Phonemes	14
The nature of symbolic representations	16
Phonetic transcription and linguistically relevant information	17
Confusion over what speech sounds ‘represent’	18
The domain of speech theory	19
2 Coarticulation: the Phenomenon	21
Introduction	21
What is coarticulation?	22
Right-to-left coarticulation	24
Phonetic under-specification	26
Overlap	26
Contextual ‘accommodation’	27
Adaptation	28
Coarticulation and ‘phonetic accommodation’	29
Coarticulation or co-production?	29
Blocking and coarticulatory spread	30
The spreading model and linear causation	31
Target theory	32
Correlation between the input target specification and the articulator target	35

Non-variable targets, variable targets and management	36
The unknown input	38
3 Coarticulation Theory	40
Introduction	40
Characterising the <i>results</i> of coarticulation	41
What must a theory of coarticulation include?	49
The focus of coarticulation theory	53
Theories and models of coarticulation	55
Models of coarticulation – presuppositions	55
Coarticulation – basic theory	57
Symbolic representation and linear context	58
Technical criticism of Wickelgren’s approach	59
Lashley’s serial ordering approach	60
Öhman’s vowel carrier model	61
MacNeilage’s three-dimensional spatial targets	61
Holmes’ linearly conjoining model	62
Lindblom’s coarticulation model	64
Lindblom and Öhman compared	64
Henke’s look-ahead model	67
Intrinsic and extrinsic allophones	70
Ladefoged’s approach to allophones	71
Classical Phonetics confuses assimilation and coarticulation	72
Kim’s ‘systematic synthesis’ model and coarticulation	74
Daniloff and Hammarberg – coarticulation as feature spreading	75
Keating’s model of coarticulation	76
Window setting – target constraints and their domains	82
Coarticulatory resistance and aggression	86
Revisions of the Window Model	87
Coarticulatory resistance	89
Co-production theories	91
Coarticulation vs. co-production	92
Sources of intrinsic variation	95
Inter-segment coarticulatory influence	96
4 Speech Motor Control	99
Introduction	99
Two dimensions to speech production theory	100
The purpose of phonetic descriptions	101
Speech production models	104
Levels of abstraction	105
Articulatory Control	106

Models of speech production	107
Serial ordering – the Lashley/MacNeilage legacy	108
Ordering in Öhman’s early model of coarticulation	110
Rhythm and temporal ordering	111
Lashley’s control mechanism	111
Action Theory	112
Task Dynamics	116
Articulatory Phonology	117
Speech production theory – changing direction	118
Unifying the approach	119
Ruling out explicit segmentation	120
The Gestural Score	120
5 Speech Production: Prosody	121
Introduction	121
Spanning – an intonation example	122
Prosodic effects	123
The correlation between abstract and physical prosody	124
Linguistic distinctiveness	124
Syllables	125
Prosody – focus on segments detracts from good prosody models	129
Features for prosody vs. features for segments	131
Syllables and stress	133
Prominence and stress	136
The metrical organisation of speech	144
Stressed syllables form the basis of rhythmic structure	145
Declination	148
Prosodic Analysis	150
Prosody and parametric modelling	152
Prosodic effects on segmental articulation	153
The basis of intonation modelling	156
 Part II Speech Perception Theory	
6 Dynamic Model Building and Expressive Content	167
Introduction	167
What levels of representation are necessary?	167
Modelling and the physical world	168
Requirements of the model – feedback	169
The speech production model	171
Cognition develops models of speech processes	172

Planes	173
Details of the speech production model	174
Supervision	176
Predictive modelling – anticipatory cognitive processing	179
Expressive content in speech production	179
The dynamic nature of the waveform	181
Prosody as information bearing	183
Sources of expressive/emotive content	184
The perception of expressive content	184
The listener is active	187
Some questions that need answering	188
Pragmatics in phonetics	189
Language pathology and clinical neuroscience	192
Language teaching/learning	192
The simulation of speech – speech technology	193
7 Speech Perception and Prosody	194
Introduction	194
Consonants vs. vowels	195
The listener’s perceptual goal	196
The principle of sufficient perceptual separation	197
The perceptual task	198
The general problem of speech perception	199
Models of Speech Perception	201
Passive models – template-based models	204
Passive theories – filtering	205
Direct Perception (as opposed to ‘mediated perception’)	206
The Motor Theory of Speech Perception	207
The Analysis by Synthesis Theory	209
Hybrid models of speech perception	209
Categories	210
Direct Realism	212
Stevens’ Quantal Theory	213
The quest for invariance	213
The Associative Store Theory	213
Similarity of production and perception	216
8 Speech Perception: Production for Perception	218
Introduction	218
Ease of articulation	220
Targets – representations of which level?	221
Adaptive Variability Theory	222

Articulator ‘response’ and its relevance to perception	223
The basic units for speech – segments again	224
Functional and physical definitions – potential conflict	225
Segments – psychological reality	226
Discrete units in the cognitive world	226
Impossibility of isolating cognitive considerations from physical considerations	227
Segment ‘boundaries’ and asynchrony	228
Two possible control models	229
Uniqueness of segments as ‘targets’	229
Trubetskoy and the continuum of speech	230
‘Rule governed’ processes?	230
General theory related to phonetic theory	231
A theory of speech production must cover both phonology and phonetics	232
Do phonological or phonetic representations dominate the perceptual processes?	233

Part III Areas of Focus, Modelling and Applications

9 Applications: Cognitive and Physical Modelling	237
Introduction	237
Duality	238
The phenomenon – what is being modelled	239
Spoken language	239
Speech and hearing/perceiving disorders	240
Language teaching	240
Speech simulations	241
Current modelling	241
An old problem	242
Cognitive and physical modelling	245
Modelling the phenomenon – theory	245
Modelling the phenomenon – the empirical approach	246
Application models within theory	247
Usefulness of empirical work	248
Testing models	249
Theories available for application – linguistically oriented models	250
Cognitive and physical modelling	251
Physical to physical – motor to acoustic	252

Deficiencies of current theory	252
When more than one model is needed	254
Metatheory	254
Our proposal	255
A computational approach	255
10 Speech Technology	256
Introduction	256
Speech synthesis – how natural?	256
Naturalness	256
Adaptation and simulation	257
Speaker expectation	258
Intelligibility – state of the art	258
Naturalness in human speech	260
Variability	261
Style	263
Data structures	264
Automatic speech recognition	265
11 Second Language Acquisition	270
Introduction	270
The relationship between production and perception in SLA	272
Production unit size and feedback	273
Application of an application	274
Theoretically motivated pronunciation training techniques	274
Accent and dialect studies	276
12 Speech Disorders	278
Introduction	278
Neuroscience applications	280
The neuropsychology of speech	280
An acquisition disorder – deafness as error	281
Cognitive approaches to understanding emotive content	283
Perception	283
Cross-modal linking	284
fMRI evidence	288
A basic assumption – phonological awareness	289
Levels of representation and the metatheory	290
Phonetics related to other areas of linguistic description	291
Diagnostic applications	292

Clinical Phonetics – procedures and expert systems:	
modelling clinical assessment	294
Areas that need further specification	296
<i>Conclusion</i>	299
<i>References</i>	300
<i>Name Index</i>	311
<i>Subject Index</i>	314

List of Figures and Tables

Figures

- 3.1 Waveform of *I want to go home*. A represents an overlapping coarticulation model: symbols indicate the start of each segment. B represents a LR inertial model: targets 1 and 3 (t1 and t3) are marked, together with the period of coarticulatory inertia (c) between the first three targets. Both models are first approximations, shown for illustrative purposes, as are the degrees of coarticulation and overlap 42
- 3.2 Waveform and spectrogram of *How are you?* Notice the continuousness of the signal and the constant ‘movement’ of the formants. There is no indication of any discrete segmentation 43
- 3.3 Waveform and spectrogram of *It’s not that I decided not to go, but that I was told to stay away*. The emphasis is on the second *not* and the word *told*. In both cases there is final plosive release, though this would not have been expected – and in general the emphasised parts of the utterance are rendered slower and with more precision 45
- 3.4 Waveforms and spectrograms of *null, moll, frond, vans, once, Banff*. In *null* and *moll* LR nasalisation can clearly be seen after the onset of the vowels, and into the final voiced segments of *frond* and *vans* – so long as the vocal cord vibration persists. With *once* and *Banff* there is no obvious LR nasalisation 46
- 3.5 Waveforms and spectrograms of *man* and *mad*. Note how nasalisation extends throughout the vowel of *man* (LR and RL coarticulation) but only part way into the vowel in *mad* (LR) coarticulation. A stylised estimate of nasal airflow has been superimposed 47
- 3.6 Waveforms and spectrograms of *i[m]put* and *i[n]put*. Since the lips are involved in both the [m] and the [p] of *i[m]put* we would expect more formant bending throughout the first part of the word, than in *i[n]put* – where the [n] does not involve the lips. There is perhaps more formant bending in *i[m]put*, but certainly not none in *i[n]put* 57