PHONETIC ANALYSIS OF SPEECH CORPORA

Phonetic Analysis *of* Speech Corpora

Jonathan Harrington

WILEY-BLACKWELL

Phonetic Analysis of Speech Corpora

Jonathan Harrington

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This edition first published 2010

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to re-use the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data is available

ISBN hbk: 978-1-4051-4169-7
ISBN pbk: 978-1-4051-9957-5

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

For Sharon

Relationship between Machine Readable (MRPA) and International Phonetic Alphabet (IPA) for Australian English

MRPA	IPA	Example
Tense vowels		
i:	iː	heed
u:	ʉː	who'd
o:	ɔː	hoard
a:	ɐː	hard
@:	ɜː	heard
Lax vowels		
I	ɪ	hid
U	ʊ	hood
E	ɛ	head
O	ɔ	hod
V	ɐ	bud
A	æ	had
Diphthongs		
I@	ɪə	here
E@	eə	there
U@	ʉə	tour
ei	æɪ	hay
ai	ɐɪ	high
au	æʉ	how
oi	ɔɪ	boy
ou	ɔʉ	hoe
Schwa		
@	ə	the

Consonants

p	p	pie
b	b	buy
t	t	tie
d	d	die
k	k	cut
g	g	go
tS	tʃ	church
dZ	dʒ	judge
H	ʰ	(Aspiration/stop release)
m	m	my
n	n	no
N	ŋ	sing
f	f	fan
v	v	van
T	θ	think
D	ð	the
s	s	see
z	z	zoo
S	ʃ	shoe
Z	ʒ	beige
h	h	he
r	ɹ	road
w	w	we
l	l	long
j	j	yes

Relationship between Machine Readable (MRPA) and International Phonetic Alphabet (IPA) for German

The MRPA for German is in accordance with SAMPA (Wells 1997), the speech assessment methods phonetic alphabet.

MRPA	IPA	Example
Tense vowels and diphthongs		
2:	ø:	Söhne
2:6	øɐ	stört
a:	a:	Strafe, Lahm
a:6	a:ɐ	Haar
e:	e:	geht
E:	ɛ:	Mädchen
E:6	ɛ:ɐ	fährt
e:6	e:ɐ	werden
i:	i:	Liebe
i:6	i:ɐ	Bier
o:	o:	Sohn
o:6	o:ɐ	vor
u:	u:	tun
u:6	u:ɐ	Uhr
y:	y:	kühl
y:6	y:ɐ	natürlich
aI	aɪ	mein
aU	aʊ	Haus
OY	ɔʏ	Beute
Lax vowels and diphthongs		
U	ʊ	Mund
9	œ	zwölf
a	a	nass

a6	aɐ	Mark
E	ɛ	Mensch
E6	ɛɐ	Lärm
I	ɪ	finden
I6	ɪɐ	wirklich
O	ɔ	kommt
O6	ɔɐ	dort
U6	ʊɐ	durch
Y	ʏ	Glück
Y6	ʏɐ	würde
6	ɐ	Vater

Consonants

p	p	Panne
b	b	Baum
t	t	Tanne
d	d	Daumen
k	k	kahl
g	g	Gaumen
pf	pf	Pfeffer
ts	ts	Zahn
tS	tʃ	Cello
dZ	ʤ	Job
Q	ʔ	(Glottal stop)
h	ʰ	(Aspiration)
m	m	Miene
n	n	nehmen
N	ŋ	lang
f	f	friedlich
v	v	weg
s	s	lassen
z	z	lesen
S	ʃ	schauen
Z	ʒ	Genie
C	ç	riechen
x	x	Buch, lachen
h	h	hoch
r	r, ʁ	Regen
l	l	lang
j	j	jemand

Downloadable Speech Databases Used in this Book

Database name	*Description*	*Language/dialect*	*n*	*S*	*Signal files*	*Annotations*	*Source*
aetobi	A fragment of the AE-TOBI database: read and spontaneous speech	American English	17	Various	Audio	Word, tonal, break	Beckman et al. (2005); Pitrelli et al. (1994); Silverman et al. (1992)
ae	Read sentences	Australian English	7	1M	Audio, spectra, formants	Prosodic, word, phonetic, tonal	Millar et al. (1994); Millar et al. (1997)
andos1	Read sentences	Australian English	200	2M	Audio, formants	Same as **ae**	Millar et al. (1994); Millar et al. (1997)
ema5 (ema)	Read sentences	Standard German	20	1F	Audio, EMA	Word, phonetic, tongue-tip, tongue-body	Bombien et al. (2007)
epgassim	Isolated words	Australian English	60	1F	Audio, EPG	Word, phonetic	Stephenson (2003); Stephenson and Harrington (2002)
epgcoutts	Read speech	Australian English	2	1F	Audio, EPG	Word	Passage from Hewlett and Shockey (1992)
epgdorsal	Isolated words	German	45	1M	Audio, EPG, formants	Word, phonetic	Ambrazaitis and John (2004)
epgpolish	Read sentences	Polish	40	1M	Audio, EPG	Word, phonetic	Guzik and Harrington (2007)

first	Five utterances from **gerplosives**						
gerplosives	Isolated words in carrier sentence	German	72	1M	Audio, spectra	Phonetic	Unpublished
gt	Continuous speech	German	9	Various	Audio, f0	Word, break, tone	Utterances from various sources
kielread	Read sentences	German	200	1M, 1F	Audio, formants	Word, phonemic, phonetic	Simpson (1998); Simpson et al. (1997)
mora	Read	Japanese	1	1F	Audio	Phonetic	Unpublished
second	Two speakers from **gerplosives**						
stops	Isolated words in carrier sentence	German	470	3M, 4F	Audio, formants	Phonetic	Unpublished
timetable	Timetable enquiries	German	5	1M	Audio	Phonetic	As **kielread**

Preface

In undergraduate courses that include phonetics, students typically acquire both skills in ear-training and an understanding of the acoustic, physiological, and perceptual characteristics of speech sounds. But there is usually less opportunity to test this knowledge on sizeable quantities of speech data, partly because putting together any database that is sufficient in extent to be able to address non-trivial questions in phonetics is very time-consuming. In the last 10 years, this issue has been offset somewhat by the rapid growth of national and international speech corpora, which has been driven principally by the needs of speech technology. But there is still usually a big gap between the knowledge acquired in phonetics from classes on the one hand and applying this knowledge to available speech corpora with the aim of solving different kinds of theoretical problems on the other. The difficulty stems not just from getting the right data out of the corpus but also from deciding what kinds of graphical and quantitative techniques are available and appropriate for the problem that is to be solved. So one of the main reasons for writing this book is a pedagogical one: it is to bridge this gap between recently acquired knowledge of experimental phonetics on the one hand and practice with quantitative data analysis on the other. The need to bridge this gap is sometimes most acutely felt when embarking for the first time on a larger-scale project, such as an honors or master's thesis, in which students collect and analyze their own speech data. But in writing this book, I also have a research audience in mind. In recent years, it has become apparent that quantitative techniques have played an increasingly important role in various branches of linguistics, in particular in laboratory phonology and sociophonetics, which sometimes depend on sizeable quantities of speech data labeled at various levels (see e.g., Bod et al. 2003 for a similar view).

This book is something of a departure from most other textbooks on phonetics in at least two ways. Firstly, and as the preceding paragraphs have suggested, I will assume a basic grasp of auditory and acoustic phonetics: that is, I will assume that the reader is familiar with basic terminology in the speech sciences, knows about the international phonetic alphabet, can transcribe speech at broad and narrow levels of detail, and has a working knowledge of basic acoustic principles such as the source-filter theory of speech production. All of this has been covered many times in

various excellent phonetics texts, and the material in Clark et al. (2007), Johnson (2004), and Ladefoged (1962), for instance, provides a firm grounding for such issues that are dealt with in this book. The second way in which this book is somewhat different from others is that it is more of a workbook than a textbook. This is partly again for pedagogical reasons: it is all very well being told (or reading) certain supposed facts about the nature of speech but until you get your hands on real data and test them, they tend to mean very little (and may even be untrue!). So it is for this reason that I have tried to convey something of the sense of data exploration using existing speech corpora, supported where appropriate by exercises. From this point of view, this book is similar in approach to Baayen (in press) and Johnson (2008) who also take a workbook approach based on data exploration and whose analyses are, like those of this book, based on the R computing and programming environment. But this book is also quite different from Baayen (in press) and Johnson (2008), whose main concerns are with statistics whereas mine are with techniques. So our approaches are complementary, especially since they all take place in the same programming environment: thus the reader can apply the statistical analyses that are discussed by these authors to many of the data analyses, both acoustic and physiological, that are presented at various stages in this book.

I am also in agreement with Baayen and Johnson about why R is such a good environment for carrying out data exploration of speech: firstly, it is free; secondly, it provides excellent graphical facilities; thirdly, it has almost every kind of statistical test that a speech researcher is likely to need, all the more so since R is open-source and is used in many other disciplines beyond speech, such as economics, medicine, and various other branches of science. Beyond this, R is flexible in allowing the user to write and adapt scripts to whatever kind of analysis is needed, it is very well adapted to manipulating combinations of numerical and symbolic data (and is therefore ideal for a field such as phonetics which is concerned with relating signals to symbols).

Another reason for situating the present book in the R programming environment is that those who have worked on, and contributed to, the Emu speech database project have developed a library of R routines that are customized for various kinds of speech analysis. This development has been ongoing for about 20 years now,[1] since the time in the late 1980s when Gordon Watson suggested to me during my post-doctoral time at the Centre for Speech Technology Research, Edinburgh University that the S programming environment, a forerunner of R, might be just what we were looking for in querying and analyzing speech data and, indeed, one or two of the functions that he wrote then, such as the routine for plotting ellipses, are still used today.

I would like to thank a number of people who have made writing this book possible. Firstly, there are all of those who have contributed to the development of the Emu speech database system in the last 20 years. Foremost Steve Cassidy who was responsible for the query language and the object-oriented implementation that underlies much of the Emu code in the R library; Andrew McVeigh who first implemented a hierarchical system that was also used by Janet Fletcher in a timing analysis of a speech corpus (Fletcher & McVeigh 1991); Catherine Watson who wrote many of the routines for spectral analysis in the 1990s; Michel Scheffers and Lasse Bombien who were together responsible for the adaptation of the **xassp** speech signal processing system[2] to Emu; and Tina John who has in recent years contributed

extensively to the various graphical user interfaces, to the development of the Emu database tool and to Emu-to-Praat conversion routines. Secondly, a number of people have provided feedback on using Emu and the Emu-R system, or on earlier drafts of this book, as well as supplied data for some of the corpora – these include most of the above and also Stefan Baumann, Mary Beckman, Bruce Birch, Felicity Cox, Karen Croot, Christoph Draxler, Yuuki Era, Martine Grice, Christian Gruttauer, Phil Hoole, Marion Jaeger, Klaus Jänsch, Pat Keating, Felicitas Kleber, Claudia Kuzla, Friedrich Leisch, Janine Lilienthal, Katalin Mády, Stefania Marin, Jeanette McGregor, Christine Mooshammer, Doris Mücke, Sallyanne Palethorpe, Marianne Pouplier, Tamara Rathcke, Uwe Reichel, Ulrich Reubold, Elliot Saltzman, Michel Scheffers, Florian Schiel, Lisa Stephenson, Marija Tabain, Hans Tillmann, Nils Ülzmann, and Briony Williams. I am also especially grateful to the numerous students both at the IPS, Munich and at the IPdS, Kiel for many useful comments in teaching Emu-R over the last seven years. I would also like to thank Danielle Descoteaux and Julia Kirk of Wiley-Blackwell for their encouragement and assistance in seeing the production of this book completed, Leah Morin for her extensive help in matters associated with copy-editing, and the very many helpful comments from four anonymous reviewers on an earlier version of this book, Sallyanne Palethorpe for her detailed comments in the final stages of writing, and Tina John both for contributing material for the online appendices and for producing many of the figures in the earlier chapters.

Notes

[1] For example in reverse chronological order: Bombien et al. (2006), Harrington et al. (2003), Cassidy (2002), Cassidy and Harrington (2001), Cassidy (1999), Cassidy and Bird (2000), Cassidy et al. (2000), Cassidy and Harrington (1996), Harrington et al. (1993), McVeigh and Harrington (1992).

[2] www.ipds.uni-kiel.de/forschung/xassp.de.html

Notes on Downloading Software

Both R and Emu run on Linux, Mac OS X, and Windows platforms. In order to run the various commands in this book, the reader needs to download and install software as follows.

I. Emu

1 Download the latest release of the Emu speech database system from the download section at http://emu.sourceforge.net.

2 Install the Emu speech database system by executing the downloaded file and following the on-screen instructions.

II. R

3 Download the R programming language from www.cran.r-project.org.

4 Install the R programming language by executing the downloaded file and following the on-screen instructions.

III. Emu-R

5 Start up R.

6 Enter `install.packages("emu")` after the `>` prompt to install the package. (You will only need to do this once.)

7 Enter `library(emu)` to load the installed package for the session.

8 Enter `emulink()` to link the package to the Emu installation and follow the instructions.

IV. Getting started with Emu

9 Start the Emu speech database tool.

- Windows: choose **Emu Speech Database System** → **Emu** from the Start menu.
- Linux: choose **Emu Speech Database System** from the Applications menu or type Emu in the terminal window.
- Mac OS X: start Emu in the Applications folder.

V. Additional software

10 Praat

- Download Praat from www.praat.org.
- To install Praat, follow the instruction at the download page.

11 WaveSurfer, which is included in the Emu setup and installed in these locations:

- Windows: **`EmuXX/bin`**
- Linux: **`/usr/local/bin`**; **`/home/'username'/Emu/bin`**
- Mac OS X: **`Applications/Emu.app/Contents/bin`**

VI. Problems

12 See FAQ at http://emu.sourceforge.net.

Contents