

Helena Spilková

**Phonetic reduction in
spontaneous speech:
an investigation of native
and non-native production**

Thesis for the degree of Philosophiae Doctor

Trondheim, January 2014

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature



NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU

Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Humanities

Department of Language and Literature

© Helena Spilková

ISBN 978-82-471-4945-4 (printed ver.)

ISBN 978-82-471-4947-8 (electronic ver.)

ISSN 1503-8181

Doctoral theses at NTNU, 2014:17

Printed by NTNU-trykk

Abstract

The aim of this thesis is to investigate the patterns of phonetic reduction (i.e. phenomena resulting from a decrease of articulatory effort) in spontaneous non-native production. To address this issue, two studies are carried out describing changes in selected phonetic parameters across domains typically associated with varying degrees of phonetic reduction. The speech materials recorded for the purposes of the present investigation included primarily spontaneous task-based dialogues in Czech, Norwegian, and English spoken by native and non-native (Czech and Norwegian) speakers. In addition, read speech in English produced by the three speaker groups with different native language (L1) background was recorded.

The first study aimed to describe the effect of speaking style (read vs. spontaneous speech) on temporal and spectral properties of the English function words *in*, *of* and *to* in productions of native speakers and two groups of non-native speakers (Czech and Norwegian speakers). The results showed that many of the phonetic parameters observed in the function words were significantly influenced by speakers' L1 background. The patterns of L1 effect, however, differed between the three observed words. While for the preposition *in*, none of the observed variables were found to vary depending on speakers' L1, in the words *of* and *to*, longer word durations in (some) non-native speakers' productions, and longer vowel durations as well as more peripheral vowel qualities in word tokens produced by non-natives as compared to native productions were revealed. This pattern of results seems to indicate non-natives' insufficient awareness of the so-called weak form words (*of* and *to*) resulting in their insufficient reduction. Further, a number of phonetic parameters were found to vary

between the two investigated speaking styles, consistently across the speaker groups. However, only some of these parameters may be reliably related to a varying degree of production effort and precision. For example, shorter word durations, higher proportions of vowel and higher proportions of voicing within the fricative in the preposition *of*, or relatively shorter durations of plosive closure in the preposition *to*, in spontaneous as compared to read speech, may be seen as lenitions resulting from a decrease of articulatory effort in spontaneous speech. On the other hand, several other variables affected by speaking style were identified that are apparently related to different aspects of speaking style than production precision.

The second study investigated the reduction of repeated mentions of content words in the course of a dialogue comparing native productions (in Czech, English and Norwegian) and productions in non-native English spoken by Czech and Norwegian speakers. The study focussed on durational, rhythmical and spectral aspects of reduction in the observed words. The results showed a consistent durational as well as spectral reduction of the repeated mentions of content words in all L1 productions, as well as in the English spoken by non-native speakers. However, a deviating pattern in the English productions of the Czech speaker group was revealed in the analysis of a rhythm-related measure (expressing the ratio of unstressed and stressed syllable durations). Here, the Czech speakers produced unusually long unstressed syllables resulting in noticeably higher unstressed-to-stressed syllable duration ratios in the first mentions of content words. In contrast to that, the ratio of unstressed and stressed syllable durations did not change much between the first mention and repeated mentions in the productions of natives and Norwegian speakers of English. This peculiar rhythmical pattern of hyperarticulation of first mentions of polysyllabic words is likely due to substantial differences in the phonological properties related to rhythm type between Czech on one side, and English and Norwegian on the other.

Preface

This thesis is submitted to the Norwegian University of Science and Technology (NTNU) for partial fulfilment of the requirements for the degree of Philosophiae Doctor.

The work referred to has been performed at the (former) Department of Language and Communication Studies, NTNU, Trondheim, under the supervision of Professor Wim van Dommelen.

The project was funded by the EU Marie Curie Research Training Network *Sound to Sense* (MRTN-CT-2006-035561).

Acknowledgements

In the first place, I would like to express my gratitude to my supervisor, Professor Wim van Dommelen. I am thankful not only for his skilled advice and the inspiration he provided at all stages of my PhD, but also for his great enthusiasm, endless patience and for encouraging me whenever I had doubts about my work. His steady support, as well as the insights about Norway and life in general that he shared with me during our tea-time sessions, goes far beyond the scope of the most thorough supervision.

This research project would not have been possible without the generous support and funding provided by the EU MC RTN Sound to Sense. Big thanks go to the Sound to Sense senior scientists for the effort invested in organising the excellent S2S workshops, and for being helpful to the fellows. I am particularly indebted to: Professor Sarah Hawkins, for being the driving force of the whole project, Professor Jacques Koreman, for always finding time to help in spite of his busy schedule, Professor Sven Mattys for providing all possible support (and organising barbecues) during my stay in Bristol, and Dr Mirjam Ernestus for initiating and largely supervising the work on the transcription and automatic alignment of the Kachna corpus. Likewise, I would like to thank all the fellows who contributed to the Kachna corpus transcription, and naturally also the speakers who were willing to provide the valuable speech material. Special thanks go to the S2S fellows for being such a great team and helping each other like good friends. I am especially thankful to Annett and Michele for their tutorials and advice on statistics, as well as for helping me struggle with my stage fright. I would also like to thank Olesya for her hospitality during my stay in Bristol, and for demonstrating that (at least for some people) everything is possible.

Special thanks are also due to Jan Volín, Radek Skarnitzl, Zdena Palková and other staff members at the Institute of Phonetics at Charles University in Prague. I am very grateful for the opportunity to use some of their recordings, but also for their help with a variety of problems, and the support I was given during the collection of my recordings in Prague. It should be mentioned that the high standards of their work and their passion for phonetic research remain an inspiration for me even in places far away from Prague.

Furthermore, I wish to thank the colleagues at the Department of Language and Communication Studies at NTNU for their help and friendly words along the way. In particular, I thank Snefrid for sharing her experience and cheering me up when it was necessary, Aleksander for his pleasant company during the years we shared the office and for always being willing to help, and Egil for å være den eneste som alltid har snakket norsk med meg. It is not possible to mention all the people involved at NTNU that I am grateful to for contributing to the smooth progress of my PhD, or for simply making my life easier.

I would also like to thank Megan Eymann Smålø for proofreading the thesis and thus helping conceal the fact that Czech has no articles. For small advice on various topics and big encouragements, I thank the master of scientific networking, Václav Jonáš Podlipský.

A much appreciated source of support on this challenging mission were my parents, relatives and friends, who either did not mind visiting me at the “end of the world” or at least did not complain about meeting much less often than desired. I would also like to thank my Czech friends in Trondheim for their cheerful company during some of the long dark evenings, for sharing the imported meat and watching Major Zeman with me. In the last year, I appreciated the help of Sweetie, who always slept on the right articles and reminded me about milk time.

My biggest thanks and sweet kisses go to Uli, who accompanied me every step of the way. I would never have made it without you.

Contents

Abstract	iii
Preface	v
Acknowledgements	vii
Contents	ix
List of figures	xiii
List of tables	xix
Abbreviations and conventions	xxiii
1 Introduction	1
1.1 Speaking styles and speech material collection	4
1.1.1 Classification of speaking styles	4
1.1.2 Types of material used in speech research	5
1.1.3 Selecting the method of speech material collection	10
1.2 Literature overview	10
1.2.1 Reduction in spontaneous conversational speech	10
1.2.1.1 Factors relevant to phonetic reduction	13
1.2.1.2 Speaking style effects	17
1.2.1.3 Repeated occurrence of words within a discourse	27
1.2.2 Relevant issues in non-native speech production	36
1.3 Selected characteristics of the studied languages	51
1.3.1 Rhythm type and related phonological properties	52
1.3.2 Connected speech processes	54
1.3.3 Voicing contrast and voicing assimilation	57
1.4 Aims and hypotheses	59
1.4.1 Aims	59

1.4.2	Hypotheses	60
1.5	Thesis outline	62
2	Speech materials	64
2.1	Task-elicited spontaneous speech	64
2.2	The speakers	65
2.2.1	Selection of non-native speakers of English.....	67
2.2.2	Assessment of non-native speakers' fluency.....	68
2.3	Spontaneous speech recordings	73
2.3.1	Elicitation task and instructions.....	73
2.3.2	Recording sessions	75
2.3.3	Audio recording and processing	76
2.4	Read speech recordings	78
3	Realisations of English function words in read and spontaneous speech.....	80
3.1	Introduction.....	80
3.2	Method	81
3.2.1	Speech material.....	81
3.2.2	Selected items	81
3.2.3	Acoustic analysis and additional observations	82
3.2.3.1	Annotations for the word in	83
3.2.3.2	Annotations for the word of	85
3.2.3.3	Annotations for the word to	86
3.2.3.4	Articulation rate estimate	87
3.2.3.5	Acoustic measurements.....	89
3.2.4	Praat scripts	90
3.2.5	Statistical analyses	90
3.3	Results.....	91
3.3.1	Context influence.....	91
3.3.1.1	Articulation rate	91
3.3.1.2	Segmental context	98
3.3.2	IN	101
3.3.2.1	Temporal organisation in the word in	101
3.3.2.2	Vowel nasalisation in the word in	103
3.3.2.3	Interaction of the word in with context	103
3.3.3	OF	110
3.3.3.1	Temporal organisation in the word of	110
3.3.3.2	Spectral measures in the word of	114
3.3.3.3	Influence of context on the realisations of the word of	118
3.3.4	TO	125

3.3.4.1	Temporal organisation in the word to	125
3.3.4.2	Spectral measures in the word to	128
3.3.4.3	Influence of context on the realisations of the word to	132
3.4	Summary and discussion	137
3.4.1	Context influences	138
3.4.1.1	Effects of neighbouring segment type.....	138
3.4.1.2	Other context effects	141
3.4.2	Temporal organisation.....	145
3.4.3	Spectral properties of speech sounds.....	155
4	Reduction of repeated mentions of content words	164
4.1	Introduction.....	164
4.2	Method.....	164
4.2.1	Speech material.....	164
4.2.2	Selected items.....	165
4.2.3	Acoustic analysis and additional observations	167
4.2.3.1	Annotations in content words.....	167
4.2.3.2	Acoustic measurements and calculating derived measures.....	168
4.2.3.3	Control factors: discourse status and prosodic structure.....	172
4.2.4	Praat scripts	174
4.2.5	Statistical analyses.....	174
4.3	Results.....	175
4.3.1	Effect of repeated mention and speaker group	175
4.3.1.1	Articulation rate and durational measures.....	175
4.3.1.2	Durational reduction of repeated mentions	177
4.3.1.3	Rhythmical aspects of reduction of repeated mentions.....	184
4.3.1.4	Spectral reduction of repeated mentions	195
4.3.2	Discourse-related control factor <i>Dialogue firstness</i>	200
4.3.3	Control factors related to prosodic structure	209
4.4	Summary and discussion	216
4.4.1	Effect of repeated mention and speaker group	217
4.4.2	Control factors	221
5	Summary and conclusions	226
5.1	Summary.....	226
5.1.1	Speech material collection.....	226
5.1.2	Realisations of English function words in read and spontaneous speech ..	227
5.1.3	Reduction of repeated mentions of content words	230
5.2	Conclusions.....	232
5.2.1	Reduction tendencies in non-native production	233

5.2.2	Deviating patterns.....	235
5.3	Methodological issues and suggestions	237
5.3.1	Future research directions.....	239
6	References.....	241
Appendix A.....	255
Appendix B.....	259
Appendix C.....	261
Appendix D.....	264
Appendix E.....	266
Appendix F.....	277

List of figures

Figure 2.1: Mean length of run (in syllables) for the two groups of non-native speakers of English with different L1 backgrounds (CZE=Czech, NOR=Norwegian).	71
Figure 2.2: Relationship between articulation rate (syll/s) and the time span between the start of English instruction and the time of recording (years) for the two groups of non-native speakers of English with different L1 backgrounds (CZE=Czech, NOR=Norwegian).	73
Figure 2.3: The positions of speakers during recording in the studios in Trondheim (left), Prague (middle) and Bristol (right). The sketches do not illustrate the dimensions of the studios (they were not the same, see text).	78
Figure 3.1: An example of segmentation and other annotations labelled in the preposition <i>in</i> (for details see text). Tier 1 shows the orthographic transcription of the displayed speech fragment.	84
Figure 3.2: An example of segmentation and other annotations labelled in the preposition <i>of</i> (for details see text). Tier 1 shows the orthographic transcription of the displayed speech fragment.	86
Figure 3.3: An example of segmentation and other annotations labelled in the preposition <i>to</i> (for details see text). Tier 1 shows the orthographic transcription of the displayed speech fragment.	87
Figure 3.4: An example of annotation of speech activity (on tier 3) and marks for syllable nuclei (on tier 4) (for details see text). Tier 1 shows the orthographic transcription of the displayed speech fragment.	88
Figure 3.5: Articulation rates (in syllables per second), for speaker groups with different L1 backgrounds (CZE=Czech, ENG=English, NOR=Norwegian) in the two speaking styles.	92
Figure 3.6: Variation coefficients (in %) of articulation rates for each speaker's read and spontaneous items. Speakers' initials are sorted alphabetically.	94
Figure 3.7: Relationship of articulation rate (in syll/s) and item duration (in ms) in the realisations of the function word <i>in</i> ($r = -0.479$; $p < 0.001$; $b = -15.33$).	95
Figure 3.8: Relationship of articulation rate (in syll/s) and item duration (in ms) in the realisations of the function word <i>of</i> ($r = -0.367$; $p < 0.001$; $b = -14.66$).	95
Figure 3.9: Relationship of articulation rate (in syll/s) and item duration (in ms) in the realisations of the function word <i>to</i> ($r = -0.495$; $p < 0.001$; $b = -18.95$).	96

Figure 3.10 (on the left) and Figure 3.11 (on the right): Distributions of left (fig. 3.10) and right (fig. 3.11) segment types (in numbers of tokens) in items pooled across the three function words, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).....	99
Figure 3.12 (on the left) and Figure 3.13 (on the right): Distributions of left (fig. 3.12) and right (fig. 3.13) segment types (in numbers of tokens) in items pooled across speaker L1 background and speaking style, for the three function words.	100
Figure 3.14: Normalised word durations (in ms) in the word <i>in</i> , for speaker groups with different L1 backgrounds (CZE=Czech, ENG=English, NOR=Norwegian) in the two speaking styles.	102
Figure 3.15: Mean values of nasalisation degree (based on auditory evaluation on a scale from 1 to 3) of the segment following the preposition <i>in</i> , for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).	106
Figure 3.16: Normalised duration (in ms) of the preposition <i>in</i> , depending on the phonological identity of the segment preceding the preposition as nasal or non-nasal, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).	107
Figure 3.17: Proportion of vowel duration (in %) within the preposition <i>in</i> , depending on the phonological identity of the segment preceding the preposition as nasal or non-nasal, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).....	109
Figure 3.18: Mean values of nasalisation degree (based on auditory evaluation on a scale from 1 to 3) of the vowel in the preposition <i>in</i> , depending on the phonological identity of the segment preceding the preposition as nasal or non-nasal, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).	109
Figure 3.19: Normalised word durations (in ms) in the word <i>of</i> , for speaker groups with different L1 backgrounds (CZE=Czech, ENG=English, NOR=Norwegian) in the two speaking styles.	111
Figure 3.20: Means of vowel proportions within the item and proportions of voicing within the fricative (in %) in the word <i>of</i> , for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).	112
Figure 3.21: Mean Bark distances of formants (F1 - F0, F3 - F2) in the vowel of the word <i>of</i> , for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).....	114
Figure 3.22: Normalised duration means (in ms) in the word <i>of</i> followed by different segment types (consonant or vowel), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).	119

Figure 3.23: Mean F1 - F0 distances (in Bark) in the vowel of the word <i>of</i> preceded by different segment types (consonant or vowel), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).	120
Figure 3.24: Proportions of voicing in the fricative (in %) in the word <i>of</i> followed by a voiceless/voiced segment, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).....	122
Figure 3.25: Distributions of phonetically voiced/voiceless segments following the word <i>of</i> (numbers of tokens), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).....	124
Figure 3.26: Normalised word durations (in ms) in the word <i>to</i> , for speaker groups with different L1 backgrounds (CZE=Czech, ENG=English, NOR=Norwegian) in the two speaking styles.	125
Figure 3.27: Means of vowel proportions within the item and proportions of release within the plosive (in %) in the word <i>to</i> , for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).	127
Figure 3.28: Mean Bark distances of formants (F1 - F0, F3 - F2) in the vowel of the word <i>to</i> , for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).....	128
Figure 3.29: Normalised duration means (in ms) in the word <i>to</i> followed by different segment types (consonant or vowel), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).	133
Figure 3.30: Mean F1 - F0 distances (in Bark) in the vowel of the word <i>to</i> followed by different segment types (consonant or vowel), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).	134
Figure 3.31: Normalised duration means (in ms) in the word <i>to</i> followed by voiceless/voiced segment, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).....	136
Figure 3.32: Distributions of phonetically voiced/voiceless segments following the word <i>to</i> (numbers of tokens), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).....	137
Figure 4.1: An example of annotations made in the observed content words (for details see text). Tier 1 shows the orthographic transcription of the displayed speech fragment.....	168
Figure 4.2: Mean durations (in ms) of the first mention and two repeated mentions of content words in Czech (CZE), English (ENG-E) and Norwegian (NOR).....	179

Figure 4.3: Mean durations (in ms) of the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).	180
Figure 4.4: Mean syllable durations (in ms) in the first mention and two repeated mentions of content words in Czech (CZE), English (ENG-E) and Norwegian (NOR).	182
Figure 4.5: Mean syllable durations (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).	183
Figure 4.6: Mean stressed syllable durations (in ms) in the first mention and two repeated mentions of content words in monosyllabic words in Czech (CZE), English (ENG-E) and Norwegian (NOR).	185
Figure 4.7: Mean stressed syllable durations (in ms) in the first mention and two repeated mentions of content words in the subset of monosyllabic words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).	186
Figure 4.8: Mean stressed syllable durations (in ms) in the first mention and two repeated mentions of content words in polysyllabic words in Czech (CZE), English (ENG-E) and Norwegian (NOR).	187
Figure 4.9: Mean stressed syllable durations (in ms) in the first mention and two repeated mentions of content words in the subset of polysyllabic words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).	188
Figure 4.10: Mean unstressed syllable durations (in ms) of the first mention and two repeated mentions of polysyllabic content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E).	190
Figure 4.11: Mean unstressed syllable durations (in ms) of the first mention and two repeated mentions of polysyllabic content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).	191
Figure 4.12: Unstressed-to-stressed syllable duration ratios in the first mention and two repeated mentions of polysyllabic content words in Czech (CZE), Norwegian (NOR) and English spoken by natives (ENG-E).	193
Figure 4.13: Unstressed-to-stressed syllable duration ratios in the first mention and two repeated mentions of polysyllabic content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).	194
Figure 4.14: Mean distance to the centroid (in Bark) for stressed vowels in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E).	197
Figure 4.15: Mean distance to the centroid (in Bark) for stressed vowels in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).	198
Figure 4.16: Mean distance to the centroid (in Bark) for vowels from each of the specified crude classes based on phonological identity, for the samples in Czech	

(CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).	199
Figure 4.17: Mean durations (in ms) in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).	202
Figure 4.18: Mean durations (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).	203
Figure 4.19: Mean syllable durations (in ms) in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).	205
Figure 4.20: Mean syllable durations (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).	206
Figure 4.21: Mean distance to the centroid (in Bark) for stressed vowels in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).	208
Figure 4.22: Mean distance to the centroid (in Bark) for stressed vowels in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).	209
Figure 4.23: Distributions of occurrence of words in final and non-final position within a higher prosodic unit (numbers of tokens) for the first mention and two repeated mentions of words in the groups based on spoken language and speakers' L1. ..	211
Figure 4.24: Word duration residuals after modelling the random effect of word identity (in ms) in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on the word's position as final or non-final in a prosodic unit.	213
Figure 4.25: Word duration residuals after modelling the random effects of word and speaker identity (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on the word's position as final or non-final in a prosodic unit.	214
Figure 4.26: Distance-to-centroid residuals after modelling the random effects of word and speaker identity (in Bark) in the first mention and two repeated mentions of	

content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on the word's position as final or non-final in a prosodic unit. 215

Figure 4.27: Distance-to-centroid residuals after modelling the random effects of word and speaker identity (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on the word's position as final or non-final in a prosodic unit. 216

List of tables

Table 1.1 – part I: Overview of the research on speaking styles.....	19
Table 1.1 – part II: Overview of the research on speaking styles	20
Table 1.1 – part III: Overview of the research on speaking styles	21
Table 1.2 – part I: Overview of the studies investigating repeated mentions of words within the discourse	28
Table 1.2 – part II: Overview of the studies investigating repeated mentions of words within the discourse	29
Table 2.1: Speaker data, including native language (L1), pair and speaker identifiers, age (years), gender (m=male, f=female), region(s) of speaker dialect, the onset of speaker’s exposure to English (years of age), and other information relevant to the speaker’s English experience.	66
Table 2.2: Selected global temporal parameters in the spontaneous English production of the non-native speakers	70
Table 2.3: Recording durations (minutes)	75
Table 2.4: Recording and audio processing specifications.....	77
Table 3.1: Means (in syll/s), standard deviations (in syll/s) and variation coefficients (in %) of articulation rates measured in speech fragments surrounding the observed words, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).....	93
Table 3.2: Overall means (in syll/s), standard deviations (in syll/s) and variation coefficients (in %) of articulation rates measured in speech fragments surrounding the observed words, for the three speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and the two speaking styles.	93
Table 3.3: Correlation coefficients and significance, regression slopes and articulation rate means (in syll/s) describing the relationship of articulation rate and item duration, for the three observed function words.....	97
Table 3.4: Numbers of tokens where the observed word is preceded by different types of segment (consonant or vowel), for the subgroups of tokens split up based on word, speaker L1 and speaking style	98

Table 3.5: Numbers of tokens where the observed word is followed by different types of segment (consonant or vowel), for the subgroups of tokens split up based on word, speaker L1 and speaking style.	98
Table 3.6: Mean F2 values (in Bark) and F2 standard deviations (in Bark) in the function word <i>of</i> , for females and males with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian).	116
Table 3.7: Mean F1 values (in Bark) and F1 standard deviations (in Bark) in the function word <i>to</i> , for females and males with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian).	129
Table 3.8: Mean F2 values (in Bark) and F2 standard deviations (in Bark) in the function word <i>to</i> , for females and males with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian).	131
Table 3.9: Overview of results of statistical analyses relating to the effects of neighbouring segment types on relevant measures of the realisation of the function words <i>in</i> , <i>of</i> and <i>to</i>	139
Table 3.10: Overview of results of statistical analyses relating to the effects of neighbouring segments' identity as nasal (for the word <i>in</i>) and neighbouring segments' phonetic voicing (for the words <i>of</i> and <i>to</i>) on relevant measures of the realisation of the function words.	142
Table 3.11: Overview of results of statistical analyses relating to variables describing the temporal organisation of the function words <i>in</i> , <i>of</i> and <i>to</i>	146
Table 3.12: Overview of results of statistical analyses relating to variables describing the spectral properties of speech sounds in the function words <i>in</i> , <i>of</i> and <i>to</i>	156
Table 4.1: Numbers of triplets (lexical items in three repetitions produced by the same speaker) selected for each combination based on spoken language and speakers' L1.	166
Table 4.2: Formant values (in Bark) of female and male centroids in Czech (CZE), British English (ENG) and Norwegian (NOR).	170
Table 4.3: Numbers and percentages of monosyllabic and polysyllabic items (triplets) in each group based on spoken language and speakers' L1.	181
Table 4.4: Numbers of triplets containing vowels from the specified crude classes based on phonological identity, and total number of triplets used for analysis of distance to the centroid, in each group based on spoken language and speakers' L1.	197
Table 4.5: Total numbers of triplets and numbers and percentages of triplets with dialogue-first status in each group based on spoken language and speakers' L1.	200
Table 4.6: Total numbers of observed word tokens and numbers and percentages of tokens evaluated as accented in each group based on spoken language and speakers' L1.	210
Table 4.7: Results of all statistical analyses concerning the effects of repeated mention (repetition) and speaker group based on spoken language or speakers' L1 background.	218
Table 4.8: Results of statistical analyses concerning the effects of repeated mention (repetition), speaker group based on spoken language or speakers' L1 background and dialogue firstness.	222

Table 4.9: Results of statistical analyses concerning the effects of repeated mention (repetition), speaker group based on spoken language or speakers' L1 background, finality and articulation rate.....	224
--	-----

Abbreviations and conventions

L1: native language

L2: second/foreign language (the distinction between these terms will not be relevant in this thesis)

C: consonant

V: vowel

F0: fundamental frequency

F1 – F3: first – third formant

VOT: voice onset time

SSBE: Standard Southern British English

Coding of statistical significance

***	$p < 0.001$
**	$0.001 \leq p < 0.01$
*	$0.01 \leq p < 0.05$
.	marginal significance ($0.05 \leq p < 0.06$)

1 Introduction

Most people would probably agree that there are only few activities that are more relaxing and enjoyable than talking with other people, regardless of whether it is an evening chat with a friend in a busy restaurant, a quick phone call with a partner, a consultation about goulash meat with a cheerful butcher or small talk at a family gathering. Most people would also not consider any of these occasions particularly difficult or demanding. The perspective dramatically changes when we consider a non-native user of the language. Talking in restaurants and at noisy parties becomes difficult and exhausting, telephone calls in a foreign language are no longer considered natural, but instead awkward or impossible. And it definitely requires a large amount of courage to ask a butcher for advice when one is not very familiar with the meat terminology in a given language.

The difficulty of communication in a foreign language (L2) has two sources. Both the perception and comprehension of speech in L2 and its production require a considerable effort. With regard to perception, understanding speech in L2 is generally considered a difficult task, and especially less proficient learners can find it very exhausting (Green, 2004). For example, non-native listeners performed worse than natives in a sentence recognition task, and the non-native disadvantage was inversely correlated with proficiency (van Wijngaarden et al., 2002). Difficulties can be expected due to missing L2 contrasts in speakers' native language (L1), which may moreover lead to the activation of inappropriate lexical items and result in more laborious speech processing (Weber and Cutler, 2004). Even for highly proficient L2 learners, L2 speech perception performance is more fragile. It has been shown that speech perception in a noisy environment or with degraded input (e.g. telephone signal) is disproportionately more difficult for non-native listeners than for native listeners (e.g. Nábělek and Donahue, 1984; Florentine et al., 1984; Mayo et al., 1997; Garcia Lecumberri and Cooke, 2006; Cutler et al., 2008).

Likewise, non-native production presents difficulties at every stage from message formulation to the production of less familiar articulatory structures. Speakers who learn a second language after a certain age typically exhibit a foreign accent that can be easily detected by native speakers of the language (Flege et al., 1995). Apart from that, non-native speech may sound ponderous and cumbersome. Although speech production involves much more than motor activity, with some simplification, a non-native speaker's production might be compared to the process of learning to swim (or learning a new swimming style). At first, the movements are strenuous and poorly coordinated, and the resulting speed is completely inadequate to the invested effort. Only after a considerable amount of practice the style is improved, the movements are perfected and swimming becomes easier and more efficient.

It is generally agreed that speech production, just as other motor activities, is governed by a principle of economy of effort (cf. Lindblom, 1990: 413-415). As a result, low-cost behaviours compatible with a given task are preferred to solutions with higher energy expenditure. For example, in informal situations with a low risk of misunderstanding, speakers tend to minimise their articulatory effort and speak less clearly than in more formal situations. Research in the last several decades has described a number of phenomena resulting from this tendency, such as the deletion of segments or syllables in casual speech (Johnson, 2004) or the spectral reduction of vowels in conversational speech compared to read text and isolated words (Koopmans-van Beinum, 1980). These phenomena are often classified under the term *phonetic reduction*. A number of studies have attempted to reveal the patterns of occurrence of reduction phenomena, and describe the factors associated with varying amount of phonetic reduction in speech. Among the investigated factors, speaking style, various measures of predictability of (parts of) the utterance, and prosodic structure have consistently been found to have a significant influence on the amount of phonetic reduction.

However, the research on phonetic reduction has so far mainly focussed on the speech production of native speakers (of various languages). Few studies have dealt with the reduction phenomena in non-native speech, or described the effects of factors relevant

Chapter 1

for phonetic reduction in non-native productions. Having mentioned the generality of motor economy, the same (native-like) tendency to minimise articulatory effort should be expected even in non-native speakers' productions. But do reduction phenomena in non-native speech occur to a degree comparable to native productions? Or could the L2 speakers' lower proficiency or pronunciation inaccuracy due to insufficient exposure to native L2 input result in a lower amount of reduction in their L2 productions? Are there any aspects that are particularly prone to interfere with some characteristics of non-native speakers' L1?

This thesis focusses on selected reduction phenomena in non-native English produced by Czech and Norwegian speakers in comparison with native productions. The thesis attempts to describe selected durational and spectral parameters across domains typically associated with varying degrees of phonetic reduction. In particular, one study deals with the form of selected English function words in read and spontaneous speech (produced by native and non-native speakers), while a second study inspects the reduction of repeated mentions of content words in the course of a dialogue (in Czech, Norwegian and both native and non-native English). A more detailed description of the occurrence of reduction phenomena in non-native speech production can contribute to the theoretical knowledge in the area of L2 production, and provide a better understanding of the interplay between a universal principle of economy of effort and language-specific reduction patterns. Results describing particular deviations of non-native reduction patterns may also be useful in L2 teaching (e.g. to help design exercises that will improve L2 learners' production in this aspect).

In order to provide a background for speech material selection we address the topic of speech materials and methods of their collection in more detail in Section 1.1. This section describes main approaches to classification of speaking styles and gives an overview of various methods that can be used to obtain recordings of different speaking styles. Section 1.2 provides a brief overview of the most important research relevant to the topics addressed in this thesis. Section 1.2.1 focusses on reduction phenomena in speech, summarising research on the factors influencing the degree of reduction

(Section 1.2.1.1), and providing a more detailed overview of the studies of speaking style effects on phonetic aspects of speech (Section 1.2.1.2), and the studies dealing with repeated occurrences of words within the discourse (Section 1.2.1.3). Section 1.2.2 then summarises the relevant issues in research on non-native speech production. Some basic phonetic and phonological characteristics of the three investigated languages (Czech, English and Norwegian), which may be relevant to the focus of this investigation, are summarised in Section 1.3. Based on the literature overview, the aim of the present investigation is further described, and general hypotheses about the outcomes of the investigation are presented in Section 1.4. An outline of the structure of the thesis is presented at the end of this chapter (Section 1.5).

1.1 Speaking styles and speech material collection

1.1.1 Classification of speaking styles

Within the area of speaking styles research, different aspects of speech have been observed with respect to their contribution to the impression of speaking style change. These include phonetic phenomena (both on the segmental and suprasegmental level) as well as lexical and syntactic parameters. At the same time, the term *speaking style* has been used in different ways by different authors. Traditionally, speaking styles research has concentrated on two types of data: spontaneous speech from unprepared situations and speech read from a prepared text. The particular types of material within these categories were then often loosely specified with regard to situation, elicitation method or speaker (e.g. interview, free conversation with a friend, sportscast) (Llisterri, 1992). Another line of speaking styles research has focussed on the distinction between casual (plain) speech and clear (hyperarticulated) speech. According to Eskénazi (1993: 502), “style reflects the action of the environment upon the individual and the individual upon the environment”, that is, speaking style is influenced by the speaker’s perception of the listener(s) and the situation, but it also reflects the speaker’s background (social level) and his wish to have a certain type or tone of conversation. The resulting speaking style is shaped by a combination of conscious and unconscious effort on the part of the speaker, and it is not always perceived in the same manner as it was intended. Based on

Chapter 1

this definition, speaking style changes may occur at any time in the course of a conversation (e.g. due to a change of conditions requiring increased intelligibility). To classify speaking styles in a systematic way, Eskénazi (1993) suggests three dimensions: the intelligibility required by the situation (and listener's needs), the familiarity between the speaker and listener(s), and the social status of the conversation participants. This definition makes it possible to place any speaking style within the three-dimensional space along the axes. For example, reading to one's own child would be characterised by a high intelligibility, a close relationship of the speaker with the listener and a relatively high cultural/social level.

Although the sophistication of the three-dimensional classification of speaking styles would allow an arbitrary selection of speaking styles for investigation, there are certain methodological restrictions for the choice of material. Barry (1995) mentions the experimental need to keep control over the investigated material (e.g. speaker parameters, discourse content, etc.), in order to guarantee that the observed differences between the materials can be attributed to the change in the speaking style. Materials consisting of read (prepared) vs. spontaneous (unprepared) speech or clear vs. casual speech can be obtained using relatively simple instructions for the speakers, which allows achieving the required control. Barry (1995) suggests that this is the reason why much speaking style research focusses mainly on the comparison of read (prepared) vs. spontaneous (unprepared) speech or clear vs. casual speech. The next section gives an overview of the various materials that can be used in speech research, and the relevant elicitation methods. This overview will provide a basis for the selection of material to be used in this thesis.

1.1.2 Types of material used in speech research

While it is relatively easy to obtain material consisting of read speech, more effort and invention has to be used to obtain spontaneous speech material. This is because we are inherently faced with the trade-off between technical quality and naturalness. Some research fields prefer the use of naturally occurring talk-in-interaction, such as recordings of telephone conversations or daily-life interactions recorded using a

portable recording device with a head-mounted microphone. There is no doubt about the naturalness and authenticity of such materials. However, the quality of such recordings is inevitably lower than the quality of recordings made in a sound-proof studio with high-quality microphones. On the other hand, we cannot talk about naturally occurring talk-in-interaction when we ask the speakers to talk together in a sound-proof studio. Still, the quality of a sound signal plays an important role in speech research, and we can find additional reasons to prefer materials that are controlled to some degree, e.g. to allow for the use of quantitative methods. This section gives an overview of the types of material used in speech research and lists some of the methods to obtain recordings of different speaking styles.

As was mentioned previously, it is not especially difficult to obtain recordings of read speech. There are many sources that are readily available (e.g. broadcast news, audiobooks...), and it is relatively easy to design texts for recording new material. Read material also provides a high control over the speech content and context. Among the materials used in speech research we can find read lists of syllables, isolated words or sentences, but also sophisticated texts specially constructed to allow the investigation of particular phenomena. In addition to read speech, there is a large variety of “laboratory” material types used for speech research with various purposes. Based on the resulting speaking style, these different types of materials can be categorised in 3 groups: elicited experimental speech, semi-spontaneous monologue and conversational speech (Jorschick, 2009).

The category of elicited experimental speech includes a number of speaking styles that do not fit in other categories. Experimental speech can be elicited using methods such as repeating utterances after the experimenter, producing a stage dialogue or recitation (after a sufficient training) or picture naming. These tasks offer a good control of the content and context of the material, but they result in less natural speaking styles. Repetition and picture naming tasks can be used in the research of speech acquisition in children (who are not able to read) and in investigations of speech impairments (Jorschick, 2009: 5).

Chapter 1

The category of semi-spontaneous monologue includes speech from a single speaker, which is elicited using a task. The task often involves the instruction to talk to another speaker (who does not participate in the recording). Semi-spontaneous monologues include narratives, interviews (where only the speech of the interviewee is collected) or description tasks for a single speaker. The resulting speech is usually quite natural, but the control over the content and context is lower than in the elicited experimental speech. An example of a monologue-eliciting description task is the description of a spatial grid-like network suggested by Levelt (1989; cited in Swerts and Collier, 1992). The speaker obtains a diagram consisting of a number of geometrical shapes in different colours arranged in a non-linear structure, and is instructed to describe the diagram so that another person can reconstruct it when listening to the recording. Swerts and Collier (1992) used these recordings for the investigation of prosodic qualities of utterances depending on their position within the discourse. Cornejo et al. (1983) described a number of other tasks suitable for eliciting spontaneous speech from children and adult speakers, such as asking speakers to provide an interpretation of a piece of art, or to give instructions about a complex process (e.g. a speaker's hobby).

Conversational speech is the most natural speaking style. There are a large number of tasks that have been used to elicit spontaneous conversational speech, including conversational tasks, object-manipulation tasks, collaborative games and discussions. Conversational tasks usually involve negotiating towards a common goal based on partly complementary information that the speakers receive, such as in the well-known Map Task (Anderson et al., 1991), where one speaker (instruction giver) has to explain a route through a map to the other speaker (instruction follower), who has a similar map where the route is not marked. Another example is the TRAINS world (Allen and Schubert, 1991; cited in Nakajima and Allen, 1993), where speakers are asked to make plans for manufacturing and shipping goods, as specified by the instructions one of them receives. Only the other speaker, however, has the knowledge of the infrastructure of the TRAINS world (a map with information about the warehouses of different goods, the factories and the train connections between them). Nakajima and Allen (1993) used

the cooperative dialogues elicited using this task for a study of prosodic patterns in relation to the discourse structure.

Another group of tasks involves manipulation with objects. Some of these object-manipulation tasks were described by Ito and Speer (2006). In one of the tasks, two participants were asked to work together to decorate a Christmas tree according to a model. This model view (on a monitor) was available to one of the speakers sitting in a soundproof booth (the instructing speaker), while the other speaker was working with the tree and a set of ornaments in a room next to the studio. The instructing speaker also had a good view of the tree that was being decorated, but he could not see the set of ornaments the decorator had available, in order to be forced to solve potential ambiguities verbally. The speakers communicated together through sets of headphones and microphones. A similar task principle is used in collaborative (computer-based) games. Some of these were described by Benus et al. (2007). In order to increase the speakers' motivation and encourage lively conversation, speakers received scoring points for successful goal accomplishment.

In discussion tasks, speakers usually get some topics that they have to talk about. However, lively discussions may be difficult to elicit just by presenting a list of topics. In the Nijmegen corpus of casual French (Torreira et al., 2010) the speakers got a list of topics including political and social issues. They were asked to choose five of these topics and discuss them in order to reach a common conclusion. Holm (2001) provided the speakers with a humorous pizza menu and instructed them to act as if they were sitting in a restaurant and discussing what kind of pizza they would prefer. The resulting recordings were used for the comparison of prosodic characteristics in spontaneous and read speech.

A special category of spontaneous conversational speech is free talk, i.e. unrestricted conversation of the participants on topics they choose themselves. Sikveland et al. (2010) describe the procedure of recording free conversations as part of the Spontal-N corpus. The authors mention the potential difficulty for the speakers to talk naturally in

Chapter 1

a studio environment, which they tried to prevent by selecting pairs of friends or acquaintances as conversation partners. The speakers had a 20-minute timeslot for free conversation, after which they could either continue for another 10 minutes, or use this time to explore and discuss the contents of a “mysterious box”. The box contained several objects that could generate curiosity and was intended to stimulate further lively conversation. Torreira et al. (2010) describe the setup of a sophisticated method to obtain recordings of free conversations where (two out of three) involved speakers were not aware of being recorded. In each session, one participant (confederate) knew about the design of the session, and was instructed to help keep a smooth flow of conversation. Using a pretext of technical problems, the confederate even left the recording studio for a period of 10 to 30 minutes, depending on the liveliness of the conversation of the two remaining participants. For ethical reasons, the participants were informed about the procedure after the end of the session, and they were asked to give their consent, and potentially add restrictions to the corpus distribution.

Besides the variety of types of laboratory speech material, it is necessary to mention the use of naturally occurring speech for research. For example, hundreds of studies in speech and speech technology research have been carried out using Switchboard (Godfrey et al., 1992), a large corpus of conversational speech, consisting of about 2500 telephone conversations by 500 volunteer speakers of American English. It is certainly a challenge to design a recording setup that enables recording daily speech interactions of good quality without an intrusive effect of the recording equipment on the speakers. Campbell (2004) describes a method of collecting natural conversational speech from volunteer speakers. The volunteers wore a portable recording device with a head-mounted microphone throughout their everyday conversational situations, providing recordings for extended periods of time. The importance of using naturally occurring talk-in-interaction in phonetic research is emphasised, for example, by Local (2003).

Within the research on casual vs. clear speech, a variety of materials were used ranging from read isolated words and other read materials to spontaneous dialogues. Moreover, different methods were used to obtain clear speech recordings. While in a number of

studies on clear speech the speakers were *instructed* to speak clearly, other studies elicited clear speech naturally by using a *communication barrier*. For instance, Hazan and Baker (2011) used two different communication barrier conditions: vocoded¹ speech and background noise (multispeaker babble), to obtain different types of naturally elicited clear speech. The advantage of this type of material is not only its greater ecological validity, but also the possibility it creates to compare the clear speech modifications in response to different barrier conditions.

1.1.3 Selecting the method of speech material collection

The examples of methods to collect different types of speech material illustrate the abundance of possibilities and show some of their advantages and disadvantages. It is obvious that different types of speech material are needed for different research goals. The choice of speech collection method for a particular study is therefore an extremely important issue. This issue, with respect to the present investigation, will be further discussed in Section 2.1 which lists the advantages and difficulties associated with using task-elicited spontaneous speech material. The rest of Chapter 2 then provides a detailed description of the speech materials used throughout this thesis.

1.2 Literature overview

1.2.1 Reduction in spontaneous conversational speech

The concept *phonetic reduction* comprises a range of phenomena that reflect deviations from canonical word forms to forms associated with less articulatory effort. The reduced forms may result from a decrease of duration or amplitude of articulatory gestures, or an increased amount of coarticulation (i.e. temporal overlap of articulatory gestures). Previous research has described various types of reduction phenomena, such

¹ Spectrally degraded speech signal simulating the hearing of a cochlear implant user

Chapter 1

as the decrease of segment duration, spectral contrast reduction², assimilations³, lenitions⁴ and segment deletions. Several recent corpus studies have also brought evidence of the frequency of occurrence of reduction phenomena in conversational speech. A study of American English showed that in conversational speech, over 20% of the word tokens had at least one segment deleted, and in more than 5% of the words, a complete syllable was deleted (Johnson, 2004). Similar deletion frequencies were described by van Bael et al. (2007) for Dutch. In their material, 20% of the words had at least one deleted segment and nearly 7% of the words lost a whole syllable. Kohler (1998) even describes the disappearance of whole words in casual German. This phenomenon is described as particularly frequent in (sequences of) function words. The completely lacking phonetic manifestation of the function word, however, does not hinder the listeners from understanding the utterance meaning, presumably due to their expectations based on context.

A plausible explanation of the occurrence of reduction phenomena is provided by the well-known Hyper- and Hypoarticulation (H&H) theory proposed by Lindblom (1990). According to this theory, speakers vary their production along a one-dimensional continuum of production precision, ranging from “hyperspeech” (very clear speech resulting from distinct articulatory gestures with minimal overlap) to “hypospeech” (reduced forms produced with minimal effort, characterised by attenuated and highly overlapped gestures). The choice of the particular phonetic form from this continuum is a product of two competing influences: (1) production economy and (2) communicative and situational demands. Since speech perception makes use of information contained in the signal as well as “signal-complementary knowledge” (e.g. context), the ideal speaker can dynamically estimate the listener’s need for explicit signal information and

² An overall decrease of the differences between realisations of sounds representing different phoneme categories and a hypothetical mean value of the system, calculated using formant values (for vowels) or cepstral features.

³ Assimilation phenomena involve a change of some aspect of a segment because of the influence of a neighbouring segment, resulting in an increased similarity of the segments.

⁴ Segment changes involving a reduction of constriction degree or an increase of segment sonority, corresponding to less articulatory effort (cf. Lass, 1984: 177-178).

adapt the production accordingly. In other words, the ideal speaker will articulate just clearly enough to ensure that the intended message can be understood in a given situation (Lindblom, 1990; 1996). A similar view on the mechanisms explaining selected reduction processes in connected speech in German was presented by Kohler (1990). According to this study, reduction phenomena result from articulatory restructuring due to the minimisation of energy expenditure, but the degree of reduction is constrained by the demands of the communicative situation.

Since there is an apparent association between a higher degree of reduction and faster speech rates, attempts have been made to describe the details of relation between articulatory precision and duration. A study by Lindblom (1963) investigated formant patterns in eight Swedish vowels embedded in three consonantal frames. The results showed a duration-dependent “vowel target undershoot” for both F1 and F2; that is, a larger deviation from the target formant values in vowels with shorter durations compared to those in longer vowels. Contrary evidence was brought by van Son and Pols (1990; 1992). They examined both the static and dynamic aspects of formant patterns in seven Dutch vowels produced in read speech at different speech rates, using speech material from one professional speaker. Apart from a higher F1 in all the vowels in fast speech, attributable to more open articulation or possibly the increased loudness of speech, the median formant values showed no difference at different speech rates (van Son and Pols, 1990). The dynamic formant tracks, inspected using 16 equidistant formant measurements per vowel, showed no large differences due to the different speech rates, either. The authors summarise that the speaker was readily able to actively adapt his articulation to a fast speech rate (van Son and Pols, 1992). A possible explanation for these results is the limited material used in these studies, consisting only of the speech of one professional speaker. This speaker, with his extensive speech training and long professional experience, was apparently capable of achieving unusual articulatory precision even at high speech rates. Due to that, the results cannot be generalised to apply to the speech behaviour of ordinary speakers in everyday situations.

Chapter 1

A number of studies investigated the articulatory activities in relation to the duration of sounds or speech rate directly. Among them, Barry (1992) focussed on the distribution of the palatalisation in consonant clusters in Russian at different speech rates. The results showed that while the distribution of palatalisation does not vary with speech rate, the electropalatographic data indicate that the size of the lingual gesture constituting palatalisation is diminished in higher speech rates. Mooshammer and Geng (2008) investigated vowel reduction in unstressed syllables in German using F1 and F2 values as well as electromagnetic midsagittal articulographic data. The results indicated that spectral reduction occurring in unstressed vowels is due to an increased amount of coarticulation with neighbouring consonants rather than due to vowel centralisation. Moreover, the results showed that the observed spectral reduction may not always be attributed to vowel shortening. In particular, lax vowels were spectrally reduced (coarticulated) without significant shortening. The authors speculate that this may be caused by a slower deactivation of muscles involved in the articulation of the neighbouring sounds, which corresponds to a lower energy expenditure.

In sum, there seems to be convincing evidence that durational shortening (e.g. due to faster speech rates) usually entails articulatory attenuation, resulting in various reduction phenomena. At the same time, other factors (e.g. increased effort in clear speech) may completely counterbalance this tendency.

1.2.1.1 Factors relevant to phonetic reduction

Exploring the causes and influential factors of phonetic reduction, just like the description of any kind of variation occurring in natural speech production, presents a considerable challenge. Previous research has shown that the occurrence of various manifestations of phonetic reduction is influenced by factors related to prosodic structure, word predictability, speech production planning and segmental context (cf. Jurafsky et al., 1998; Bell et al., 1999; Lavoie, 2002). The following paragraphs will present an overview of research that evaluated the effects of some of these factors.

There is a large amount of evidence which indicates that various aspects of prosodic structure influence the degree of reduction in speech. A number of studies on various languages have observed the effects of lexical stress as well as higher-level prosodic prominence on durational and spectral reduction. For example, spectral reduction of vowels was observed in lexically unstressed syllables as compared to stressed ones (Koopmans-van Beinum, 1980). In a corpus study by van Bael et al. (2007), phones in lexically unstressed syllables were more frequently deleted. The effects of lexical stress and pitch accent on vowel durations were described by Van Santen (1992). Van Bergem (1993) confirmed the effect of lexical stress and sentence accent, as well as word class (i.e. function word vs. content word) on syllable duration and spectral reduction of vowels. Similarly, the position of syllables relative to the pauses and boundaries of prosodic units was observed to affect their durational properties. The well-known effect of final lengthening was described in word-final, phrase-final as well as utterance-final syllables (Oller, 1973). Furthermore, Wightman et al. (1992) showed that the amount of constituent-final lengthening is proportional to the degree of the following prosodic boundary.

Besides prosodic factors, certain factors related to linguistic redundancy and predictability have also been shown to influence the degree of phonetic reduction. It was observed that less predictable elements tend to be articulated more carefully than more predictable ones (Lieberman, 1963; Fowler and Housum, 1987; Aylett and Turk, 2004). Various factors contribute to word predictability, including the word's lexical frequency and its probability in a given context. Probabilistic relations between words are often expressed using measures such as the joint probability of a target word with a neighbouring word, the conditional probability of a target word given the preceding word, or given the following word, etc. Clearly, highly predictable linguistic units can be considered more redundant in the discourse, and therefore, the effort required for their production can be minimised. This view is consistent with Lindblom's H&H theory (1990). The effect of word lexical frequency on phonetic reduction was shown, for example, by Bell et al. (2002). This corpus study examined the duration of content words in a large corpus of spontaneous American English (Switchboard; see Section

Chapter 1

1.1.2 for details) and studied the role of several factors relating to word predictability: word frequency, conditional probabilities, joint probabilities, semantic relatedness and word repetition. A number of additional factors, including disfluencies, position in a phrase, speech rate, etc., were controlled for. The results indicated that word lexical frequency is the strongest predictor of content word reduction. In addition, the conditional probability of a word's occurrence given the following word, had a significant effect on reduction, while the other investigated factors did not have any additional effects. The effect of lexical frequency on durational and spectral reduction was also shown in a study by Guion (1995). She examined four pairs of homophones with different lexical frequencies (e.g. *need* vs. *knead*) and found that while there were no differences between the homophones in citation form, in a sentence context, the more frequent homophone was more reduced than the less frequent one. In particular, F2 values in the stressed vowel were found to differ in all the investigated homophone pairs for the majority of speakers, and word duration differed in two of the investigated homophone pairs. Other probabilistic factors were found to influence the word duration and frequency of vowel reduction in a study examining the ten most frequent function words in the Switchboard corpus (Bell et al., 2003). The results showed that high conditional probabilities with previous and following words were associated with shorter durations and higher frequency of vowel reduction. Torreira and Ernestus (2009) showed that the closure duration in intervocalic /t/ in French is influenced by the word bigram frequency (i.e. the joint probability of a target word with the following word). A study by Schuppler et al. (2012) found similar results for Dutch, showing that in particular, word bigram frequency has an effect on the acoustic presence or absence of word-final /t/ in a large corpus of automatically annotated spontaneous speech. In addition, an analysis of the presence of sub-phonemic properties carried out on a smaller subset of tokens confirmed a more frequent absence of individual sub-phonemic properties, such as constriction, burst, or alveolar friction, in tokens with higher word bigram frequency. Also, the repeated occurrence of a word within the discourse is closely associated with word predictability. Since the phonetic reduction of repeated mentions of content words will be one of the topics dealt with in this thesis, studies

investigating various acoustic characteristics of repeated mentions will be discussed in more detail in Section 1.2.1.3.

Another strong predictor of word duration is related to speech production planning problems. Fox Tree and Clark (1997) documented in a large corpus of spontaneous English that the form of the English definite article *the* varies as a function of production planning problems. They found that when the article was followed by production difficulties (e.g. disfluency), its vowel was most frequently produced as non-reduced /i:/ rather than as a schwa. A study by Bell et al. (1999) investigated the effect of a number of factors on the word duration and vowel quality of the ten most frequent function words in a corpus of conversational American English (Switchboard; see Section 1.1.2 for details). The evaluated factors included the presence of disfluency, various measures of predictability, the word's position within turn and utterance, and the speaker's age and sex. The results showed that the presence of a filled pause following the word was the strongest predictor of word lengthening. Bell et al. (2003) used the Switchboard corpus to investigate the effect of predictability, disfluencies and utterance position on the variation of the function word forms in conversational English. The results relevant to the effects of disfluencies showed that all types of disfluencies (categorised as silent pause, filled pause and repetition), both preceding and following the observed function words, have a strong effect on function word forms. Words tend to be longer and have less reduced vowels in the neighbourhood of disfluencies, the effect being stronger for disfluencies following the observed word. The strength of the above described effects motivated many studies to adopt criteria for excluding materials likely to be affected by disfluencies (e.g. Bell et al., 2002; Pluymaekers et al., 2005; Baker et al., 2011).

Previous research has also found evidence of the influence of segmental context on word form variability. Several recent corpus studies of conversational American English and Dutch have shown that the type of immediately following segment (consonant or vowel) influences several measures of reduction. The results showed that words tend to be shorter, have more reduced vowels and more frequently deleted segments, when they

Chapter 1

are followed by a consonant (Jurafsky et al., 1998; Bell et al., 2003; van Bael et al., 2007).

The amount of reduction was also found to vary depending on gender and age (Byrd, 1994; Bell et al., 1999). Byrd (1994) found that male speakers reduce more than females. In the study by Bell et al. (1999), females were found to have lower speech rates, and produce less reduced forms than males. Older speakers were also found to have a lower speech rate and a lower amount of reduced forms, the age effect being even stronger for females than for males.

1.2.1.2 Speaking style effects

By definition, different speaking styles may vary in the degree of production effort and precision as a reaction to differing demands of the situation (cf. Section 1.1.1). Therefore, speaking style may be considered an important factor influencing the degree of phonetic reduction. A number of studies have been carried out using materials in different languages, and usually comparing spontaneous productions with read speech. Table 1.1 provides an overview of the research investigating the durational and spectral properties of speech sounds, the frequency of some types of reduction phenomena and the various prosodic characteristics of speech productions in different speaking styles. The table summarises the investigated languages, types of speech material used and the main findings. The following paragraphs will then present more details of the mentioned studies.

With regard to durational reduction, shorter segment durations were found in spontaneous vs. read speech in Dutch (van Son and Pols, 1999), Russian (Bondarko et al., 2003; Bolotova, 2003) and Finnish (de Silva et al., 2003). In contrast to the results above, de Silva et al. (2003) found a tendency to use shorter durations of segments in read speech for Russian and Dutch. Moreover, spontaneous speech was characterised by larger durational variability of segments (Bolotova, 2003). Furthermore, other reduction phenomena, such as consonant cluster simplifications, consonant weakening, vowel

centralisations and syllable elisions were observed more frequently in spontaneous vs. read speech in Bulgarian, Czech, Greek, Italian, Polish and Russian (Barry and Andreeva, 2001). Similarly, a larger number of elisions and assimilations of final /n/ was observed in spontaneous speech in Finnish, and a more frequent elision of /j/ was found in spontaneous Russian, while the redundant final /n/ in Dutch plural suffixes was observed more frequently in read speech (de Silva et al., 2003).

Chapter 1

Table 1.1 – part I: Overview of the research on speaking styles

	Study	Language(s)	Types of speech material	Findings	Notes
Segment durations, elisions and other processes	van Son and Pols (1999)	Dutch	spontaneous speech (prepared stories) vs. read speech (read version of the spontaneous speech transcript)	•shorter consonant durations in spontaneous speech	one professional speaker
	Bondarko et al. (2003)	Russian	spontaneous speech (dialogues) vs. read speech (read version of the spontaneous speech transcript)	•shorter segment durations in spontaneous speech	
	Bolotova (2003)	Russian	spontaneous speech (dialogues) vs. read speech (read version of the spontaneous speech transcript)	•shorter segment durations in spontaneous speech •greater segment duration variability (of vowels and sonorants) in spontaneous speech	
	de Silva et al. (2003)	Finnish	spontaneous speech (dialogues) vs. read speech (read version of the spontaneous speech transcript)	•shorter segment durations in spontaneous speech •more frequent elisions or assimilations of final /n/ in spontaneous speech	
		Russian	spontaneous speech (dialogues) vs. read speech (read version of the spontaneous speech transcript)	•longer segment durations in spontaneous speech •more frequent elision of /j/ in spontaneous speech	
		Dutch	spontaneous speech (story telling) vs. read speech	•longer segment durations in spontaneous speech •redundant final /n/ in plural suffixes more frequent in read speech	
Barry and Andreeva (2001)	Bulgarian, Czech, Greek, Italian, Polish and Russian	spontaneous speech (interactional task) vs. read speech (read version of the spontaneous speech transcript)	•consonant cluster simplifications, consonant weakening, vowel centralisations and syllable elisions more frequent in spontaneous speech		
Vowel reduction (formants)	Koopmans-van Beinum (1980)	Dutch	isolated vowels, canonical word forms, read speech, retold story, conversational speech	•vowel quality contrast decreases in more spontaneous productions	
	Harmegnies and Poch-Olivé (1992)	Spanish	spontaneous conversational speech vs. laboratory speech (i.e. word list reading)	•vowel centralisation and greater within-category scatter in spontaneous vs. laboratory speech	
	Laan (1997)	Dutch	spontaneous speech on prepared topic, read speech (read version of the spontaneous speech transcript), isolated vowels	•smaller vowel space in both speaking styles as compared to vowels produced in isolation •more centralised vowel formant values in spontaneous speech (only for one speaker)	only 2 speakers (one of them being a professional speaker)
	Bondarko et al. (2003)	Russian	spontaneous speech (dialogues) vs. read speech (read version of the spontaneous speech transcript)	•greater variability of formant values for peripheral vowels /a/, /i/ and /u/ in spontaneous speech	
	Moon and Lindblom (1994)	English	citation forms (i.e. normal reading) vs. clear speech	•less formant displacement due to context in clear speech	

Table 1.1 – part II: Overview of the research on speaking styles

	Study	Language(s)	Types of speech material	Findings	Notes
Consonant reduction, cepstral space reduction	Van Son and Pols (1999)	Dutch	spontaneous speech (prepared stories) vs. read speech (read version of the spontaneous speech transcript)	<ul style="list-style-type: none"> • lower values of centre of gravity in spontaneous speech (for all consonants except for plosives) • smaller intervocalic sound energy difference in spontaneous speech (for all consonants except for nasals) 	one professional speaker
	Furui et al. (2005)	Japanese	read speech vs. spontaneous speech (i.e. presentations, informal presentations and dialogues)	<ul style="list-style-type: none"> • reduced cepstral differences between the phonemes in spontaneous speech 	
	Nakamura et al. (2008)	Japanese	read speech, monologue, dialogue	<ul style="list-style-type: none"> • reduced cepstral differences between the phonemes in spontaneous speech • increased variance for almost all phonemes in spontaneous speech 	
	Rouas et al. (2010)	French	read speech, prepared and casual spontaneous speech	<ul style="list-style-type: none"> • reduction of spectral space and increased spectral variance of each phoneme in both types of spontaneous speech vs. read speech • more noticeable reduction in prepared spontaneous speech than in casual speech 	
Prosody	Laan (1997)	Dutch	spontaneous speech on prepared topic, read speech (read version of the spontaneous speech transcript), isolated vowels	<ul style="list-style-type: none"> • higher median F0, broader F0 range, greater F0 variation within an utterance and more frequent F0 declination throughout an utterance in read as compared to spontaneous speech (only significant for one speaker) 	2 speakers (one of them being a professional speaker)
	de Silva et al. (2003)	Dutch	spontaneous speech (story telling) vs. read speech	<ul style="list-style-type: none"> • lower mean F0 in spontaneous vs. read speech 	
		Finnish	spontaneous speech (dialogues) vs. read speech (read version of the spontaneous speech transcript)	<ul style="list-style-type: none"> • higher mean F0 in spontaneous vs. read speech 	
	Face (2003)	Spanish	spontaneous speech (various interviews)	<ul style="list-style-type: none"> • common occurrence of words with no F0 rise in their stressed syllable, less frequent use of final lowering in spontaneous speech, as compared to previous findings from well-studied lab speech 	results from spontaneous speech compared to previous findings from lab speech
	Mixdorff and Pfitzinger (2005)	German	spontaneous speech (Map Task dialogues) vs. read speech (read version of the spontaneous speech transcript)	<ul style="list-style-type: none"> • smaller number of accents, lower amplitudes of associated accent commands in spontaneous speech 	
	Howell and Kadi-Hanifi (1991)	English	spontaneous speech (description) vs. read speech (read version of the spontaneous speech transcript)	<ul style="list-style-type: none"> • different prosodic unit boundaries and stress positioning in spontaneous and read speech • fewer pauses in read speech 	
	Bondarko et al. (2003)	Russian	spontaneous speech (dialogues) vs. read speech (read version of the spontaneous speech transcript)	<ul style="list-style-type: none"> • shorter pauses in spontaneous speech • missing or very short pauses on boundaries of intonation units more frequent in spontaneous speech 	
	Cucchiari et al. (2002)	non-native productions in Dutch	read vs. spontaneous speech (different tasks)	<ul style="list-style-type: none"> • longer mean duration of silent pause in spontaneous speech • considerably higher frequency of silent pauses, filled pauses and disfluencies in spontaneous speech 	non-native speakers with different proficiency levels

Chapter 1

Table 1.1 – part III: Overview of the research on speaking styles

	Study	Language(s)	Types of speech material	Findings	Notes
Speech rate	Koopmans-Van Beinum (1992)	Dutch	spontaneous speech on prepared topic vs. read speech (read version of the spontaneous speech transcript)	<ul style="list-style-type: none"> • significantly higher articulation rate in spontaneous speech • larger speech rate range and variability in spontaneous speech 	one professional speaker
	Laan (1997)	Dutch	spontaneous speech on prepared topic, read speech (read version of the spontaneous speech transcript), isolated vowels	<ul style="list-style-type: none"> • higher articulation rate in spontaneous speech (only significant for one speaker) 	2 speakers (one of them being a professional speaker)
	Cucchiariini et al. (2002)	non-native productions in Dutch	read vs. spontaneous speech (different tasks)	<ul style="list-style-type: none"> • higher articulation rate in spontaneous speech for beginners and intermediate speakers • lower speech rate in spontaneous speech for beginners and intermediate speakers 	non-native speakers with different proficiency levels
	Hirschberg (2000)	English	spontaneous (interactional task) vs. read speech (read version of the spontaneous speech transcript)	<ul style="list-style-type: none"> • consistently lower speech rates in spontaneous speech 	
	Mixdorff and Pfitzinger (2005)	German	spontaneous speech (Map Task dialogues) vs. read speech (read version of the spontaneous speech transcript)	<ul style="list-style-type: none"> • lower perceptual local speech rate in spontaneous speech • tendency to greater speech rate variation in spontaneous speech 	

As for the spectral properties of speech sounds, previous research has focussed both on the vowel formant values and on the spectral properties of consonants. Koopmans-van Beinum (1980) investigated the quality of Dutch vowels produced in various speech conditions, based on the stress position and speaking style. To describe spectral reduction, a formant centroid⁵ value was calculated for each speaker, which allowed the measuring of the acoustic system contrast⁶ in different speech conditions. The results showed that vowel quality contrast decreased in more spontaneous productions. Similarly, Harmegnies and Poch-Olivé (1992) observed vowel centralisation and greater within-category scatter in spontaneous vs. laboratory speech (i.e. word list reading) in Spanish. They pointed out, however, that these results may stem from language-specific properties, in particular the simple vocalic system of Spanish. They assume that in languages with a rich vowel system, vowel centralisation may present a larger hindrance

⁵ The speaker-specific centroid value was calculated as an average of formant values of all Dutch vowels produced by the speaker.

⁶ The acoustic system contrast was determined using the measure of “vowel distance to the centroid” calculated as the Euclidean distance of a given vowel from the centroid.

of intelligibility. Also, a study by Laan (1997) inspected formant values, comparing isolated productions of vowels, and vowels from read and spontaneous speech in Dutch. The results showed that vowels produced in both speaking styles shifted from their ideal position, defined as formant values measured in vowels produced in isolation. The resulting vowel space in both speaking styles was smaller than the space formed by vowels produced in isolation. Moreover, one of the two speakers produced significantly more centralised vowel formant values (i.e. more divergent from the ideal formant values) in spontaneous speech than in read speech, but the vowels of the second speaker did not show significant differences between the speaking styles. Bondarko et al. (2003) also observed greater variability of the formant values for the peripheral vowels /a/, /i/ and /u/ in spontaneous vs. read speech in Russian.

A different selection of speaking styles was used in a study by Moon and Lindblom (1994). They investigated differences between words in citation forms (normal reading) and in clear speech⁷ in English. This study addressed the question of duration-dependent undershoot of the vowel targets that had been found in earlier research (e.g. Lindblom, 1963). The results of the study by Moon and Lindblom (1994) revealed systematic patterns of formant displacement in English front vowels embedded in /wVl/ sequences. Formant frequencies were found to be displaced in the direction of the frequencies of the neighbouring consonants. Consistent with the duration-dependent undershoot hypothesis, the amount of formant displacement varied with vowel duration (showing less displacement in longer durations). However, there were differences between the speaking styles: the items produced in clear speech had less displaced formants. The smaller degree of reduction in clear speech was partly due to longer vowel durations, but it was also shown that the velocity of F2 transition was higher in clear speech.

Several studies have also inspected the spectral properties of consonants in different speaking styles. Van Son and Pols (1999) investigated the acoustic properties of

⁷ The speakers were instructed to pronounce the words as clearly as they could.

Chapter 1

consonants in read and spontaneous speech in Dutch. They chose several acoustic correlates of consonant reduction, relating to articulation and speech effort, such as the spectral balance and intervocalic sound energy difference between vowels and consonant. Apart from shorter durations of consonants observed in spontaneous vs. read speech, the results showed lower values of the centre of gravity⁸ in all consonants except for plosives. Also, the measure of intervocalic sound energy difference varied between speaking styles: in all segment types except for nasals, the intervocalic sound energy difference was smaller in spontaneous vs. read speech. A study by Furui et al. (2005) used cepstral feature vectors⁹ to compare the acoustic properties of vowels and consonants in read speech with spontaneous speech using a large corpus of Japanese. The spontaneous speech in this corpus included academic presentations, informal presentations and dialogues. The results showed reduced cepstral differences between the phonemes in spontaneous speech, compared to read speech. Similar results were reported by Nakamura et al. (2008), who also observed an extension of within-category variance for almost all phonemes in spontaneous vs. read speech. Rouas et al. (2010) carried out a similar study comparing read speech with prepared and casual spontaneous speech in French. The results of this study confirmed the reduction of spectral space and increased spectral variance of each phoneme in spontaneous vs. read speech in French. The reduction effect was more noticeable in prepared spontaneous speech as compared to that found in casual speech.

Moreover, a number of studies have investigated prosodic aspects and other global characteristics of speech productions associated with the use of different speaking styles. Laan (1997) investigated several measures relating to F0 in read and spontaneous speech in Dutch. The results showed a higher median F0, a broader F0 range, a greater F0 variation within an utterance and a more frequent F0 declination throughout an

⁸ The measure of the centre of gravity represents the “mean” frequency, having lower values for spectra with strongly represented lower-band frequencies, and higher values in spectra with strong high frequencies and weaker low frequencies.

⁹ Cepstral feature vectors consist of a number of coefficients calculated from the transformed speech signal spectrum.

utterance in read as compared to spontaneous speech. It needs to be mentioned, however, that the results were only significant for one of the two speakers used in the study, while the other speaker only showed similar tendencies that were not significant. Similarly, de Silva et al. (2003) found a higher mean F0 in read vs. spontaneous speech in Dutch. On the other hand, the F0 mean was lower in read than in spontaneous speech in Finnish (de Silva et al., 2003). Face (2003) investigated several intonational features in Spanish declaratives, comparing the properties of spontaneous speech to previous findings from laboratory speech. In contrast to well-studied lab speech, where (nearly) all stressed syllables are accompanied by an F0 rise, it was not uncommon to find words without any F0 rise in their stressed syllable in spontaneous speech. Also, final lowering, which was very frequent in lab speech, proved to be rather uncommon in spontaneous speech. This phenomenon is then used almost uniquely to mark repeated or predictable information. Mixdorff and Pfitzinger (2005) compared spontaneous speech elicited using the Map Task (see Section 1.1.2 for details) and read speech obtained using transcripts of the spontaneous speech productions. The results showed only minor style-related differences in F0 means and standard deviations. The measured F0 contours were also used to automatically estimate Fujisaki parameters representing accent commands. The number of accented syllables and associated accent commands' amplitudes were found to be higher in read than in spontaneous speech.

Other prosodic features, such as stress position, phrasing and pausing patterns were also found to differ between the speaking styles. Howell and Kadi-Hanifi (1991) compared spontaneous speech productions with read speech (using the transcripts of speakers' original spontaneous productions) in English. They found that the prosodic unit boundaries and stress positioning differed between spontaneous and read speech. Moreover, they found fewer pauses in read speech. Bondarko et al. (2003) compared spontaneous and read speech in Russian. Their results showed shorter pause durations in spontaneous speech as compared to those found in read speech. In addition, it occurred more frequently in spontaneous speech that no pause, or a very short pause, was realised on boundaries of intonation units. An important contribution to speaking style research investigating *non-native* productions was brought by Cucchiari et al. (2002). In this

Chapter 1

study, Dutch productions of non-native speakers with different proficiency levels were investigated in both read and spontaneous speech. The results showed a longer mean duration of silent pauses, as well as a considerably higher frequency of silent pauses, filled pauses and disfluencies in spontaneous speech.

Another parameter describing the global characteristics of speech productions in different speaking styles is speech rate¹⁰ (in some of the studies articulation rate¹¹). The studies described below present rather contradicting results regarding speech rate differences between the speaking styles. Koopmans-Van Beinum (1992) studied a number of parameters in read and spontaneous speech in Dutch and found a significantly higher speech rate in spontaneous speech. In addition, spontaneous speech showed a larger speech rate range and variability. In Laan (1997) the articulation rate was also higher in spontaneous speech than in read speech in Dutch. This was, however, only found significant for one of the two speakers used in the study. In the above mentioned study of non-native speakers of Dutch, Cucchiari et al. (2002) found that both beginners and intermediate speakers had higher articulation rates in spontaneous production, this difference being smaller in intermediate speakers (it has to be mentioned that their spontaneous speech task was more cognitively demanding than that of beginners). In addition, the study presented data on the advanced speakers' production of read speech. Their articulation rate was higher than the rates of both (lower) proficiency groups in spontaneous speech, but the comparison controlled for proficiency level could not be made since there were no spontaneous speech data from the advanced speaker group available. On the other hand, both beginners and intermediate speakers had considerably lower speech rate values in spontaneous speech as compared to read speech. This result is apparently related to a significantly higher overall amount of pauses in spontaneous speech. Similarly, Hirschberg (2000) found consistently lower speech rates in spontaneous speech. The extent, to which this result is due to a higher amount of pauses in spontaneous speech and due to articulation rate

¹⁰ Speech rate is usually measured in syllables per second, phones per second or words per minute.

¹¹ Articulation rate is defined as the speech rate measured in the speech sample excluding pausing time.

reduction, remains to be clarified. Mixdorff and Pfitzinger (2005) also observed a higher perceptual local speech rate¹² in read speech. Moreover, spontaneous speech showed a tendency to have a greater speech rate variation. However, individual subjects were observed to use different strategies to achieve different speaking styles.

The inconsistencies in the results above may result from several factors. Apart from the fact that the different measures (i.e. speech rate vs. articulation rate) may show different tendencies (cf. results of Cucchiarini et al., 2002), differences may be expected due to the sampling of speakers used in the studies and the types of speech tasks used for the elicitation of speech. While Koopmans-Van Beinum (1992) used only one speaker, who was a professional speaker, and Laan (1997) used two speakers, one of which was a professional newsreader, the study by Hirschberg (2000) used material from 17 speakers, and Mixdorff and Pfitzinger (2005) used material from four university students. It may be unreliable to generalise from the results of studies using low numbers of speakers, or using material produced by speech professionals. Moreover, there were considerable differences in the conditions at which spontaneous speech was produced. The studies by Koopmans-Van Beinum (1992) and Laan (1997) used narratives on well-prepared topics, while Mixdorff and Pfitzinger (2005) used speech elicited using an interactional task (particularly the Map Task). Similarly, the speakers in Hirschberg's (2000) study interacted with a simulated voice response system attempting to make air travel plans. The particular task used for spontaneous speech elicitation is likely to have a major influence on various speech characteristics, and the elicitation method differences should be kept in mind whenever results of different studies are compared.

On the whole, we could observe relatively consistent tendencies in the parameters relating to spectral properties across the different studies comparing speaking styles. On the other hand, the parameters relating to prosody and temporal measures often showed

¹² This measure is based on a linear combination of local syllable rate and local phone rate.

Chapter 1

diverging results both across the studies and between the speakers within a study (e.g. Laan, 1997). Apart from reasons stemming from differences in language and material type, the inconsistencies may be due to speaker-specific strategies to achieve a given speaking style (cf. Holm, 2001: 50-51; Eskénazi, 1993: 504).

1.2.1.3 Repeated occurrence of words within a discourse

As discussed in Section 1.2.1.1, various measures relating to word predictability were shown to have an effect on the degree of phonetic reduction. A traditionally investigated factor related to word predictability is a word's status as "new" (first mention of the word) or "given" (repeated mention) within the discourse. Table 1.2 provides an overview of the studies describing the occurrence of reduction phenomena in repeated mentions of words and the consequences of the reduction on the words' intelligibility. Some of the studies have also identified certain limitations of the general tendencies and interacting factors. The table summarises the investigated languages, types of speech material used and the main findings. The following paragraphs will then present more details of the mentioned studies.

The effect of word repetition on phonetic reduction was first reported by Fowler and Housum (1987). This study using an unscripted monologue and interviews showed that repeated mentions of content words have shorter durations, lower peak amplitude in the word's stressed vowel, and are less intelligible when presented in isolation, when compared to first mentions. On the other hand, no significant differences were found in the average F0 measured in the word's stressed vowel. In addition, longer words were found to shorten more than shorter words. In a further study (Fowler, 1988), no shortening was found for repeated content words produced in lists, but the word repetitions were shortened when integrated in meaningful paragraphs read aloud. In the experiment using read meaningful texts, it was also shown that words preceded by homophones do not shorten, while words preceded by synonyms shorten slightly (the results were not conclusive). The fact that words preceded by homophones do not reduce contradicts the possibility that the reduction of repeated mentions of words is an

effect of articulatory priming. A final experiment showed a greater shortening effect in repeated words produced in a communicative context (spontaneous monologues directed to a listener), than in those used read text (transcripts of the monologues read by the same speakers). This result may, however, be partly due to a slower speech rate (and thus larger “room” for shortening) in spontaneous speech.

Table 1.2 – part I: Overview of the studies investigating repeated mentions of words within the discourse

Study	Language(s)	Types of material	Findings	Notes
Fowler and Housum (1987)	English	unscripted monologue and interviews	<ul style="list-style-type: none"> shorter durations, lower peak amplitude in the word's stressed vowel and lower intelligibility in isolation for repeated mentions of content words 	
Fowler (1988)	English	word lists, read meaningful texts, monologues	<ul style="list-style-type: none"> no shortening for repeated content words produced in lists; repeated content words were shortened when integrated in meaningful paragraphs read aloud words preceded by homophones do not shorten, words preceded by synonyms shorten slightly greater shortening effect in repeated words produced in a communicative context (spontaneous monologues directed to a listener) 	
Koopmans-van Beinum, van Bergem (1989)	Dutch	spontaneous and read speech, isolated words	<ul style="list-style-type: none"> no effect of word repetition on either of the inspected parameters in read speech F0 and distance to the centroid differed significantly between first and repeated mentions 	one professional speaker
Shields and Balota (1991)	English	sentences produced from memory	<ul style="list-style-type: none"> shortest word durations in word repetitions within a sentence associatively related words shorter than words unrelated to a previous word in the sentence lower peak amplitude in word repetitions 	
Hawkins and Warren (1994)	English	spontaneous conversations	<ul style="list-style-type: none"> no significant effect of word repetition on word intelligibility, when pitch accent is controlled for 	
McAllister et al. (1994)	English	problem-solving task in monologue and in interaction with another speaker	<ul style="list-style-type: none"> overall shorter word durations in dialogues vs. monologues significantly reduced second mentions of words in dialogues (but not in monologues) 	
Fisher and Tokura (1995)	English	verbal descriptions of short puppet events	<ul style="list-style-type: none"> shorter durations, lower relative amplitude, lower pitch and less pitch variability in repeated mentions of target words 	mothers talking to their infants
Fowler et al. (1997)	English	spontaneous narrations of a movie content	<ul style="list-style-type: none"> duration shortening of repeated occurrences of words within but not between episodes 	
Gregory et al. (1999)	English	Switchboard corpus subset (telephone conversations)	<ul style="list-style-type: none"> number of preceding mentions of a word within a discourse and a word's semantic relatedness to preceding discourse contribute to word shortening 	
Bard et al. (2000)	English	series of Map Task dialogues	<ul style="list-style-type: none"> word intelligibility decrease on second mention even when the word's first occurrence was directed to a different follower, or when the listener reported inability to see the respective landmark shortening and intelligibility decrease of a word's second mention takes place regardless of which participant produced the first mention 	each instruction giver described a route through the same map in two sessions with different instruction followers
Aylett and Turk (2004)	English	Map Task dialogues	<ul style="list-style-type: none"> strong effect of several redundancy factors (including the number of previous mentions of a landmark) as well as prosodic factors on syllable durations 	

Chapter 1

Table 1.2 – part II: Overview of the studies investigating repeated mentions of words within the discourse

Study	Language(s)	Types of material	Findings	Notes
Plymaekers et al. (2005)	Dutch	face-to-face conversations	<ul style="list-style-type: none"> • number of previous mentions of a word influences duration of word suffix, independent of pitch accent influence • word stem duration and number of segments only influenced by word predictability from neighbouring words 	
Trón (2008)	English	Map Task dialogues	<ul style="list-style-type: none"> • durational reduction between consecutive mentions of words largest at short latencies (time elapsed between the word mentions); no significant reduction at latencies over 60 seconds • larger reduction between early mentions of a word in a dialogue (e.g. first and second mention) 	
Baker and Bradlow (2009)	English	read paragraphs in plain and clear speaking style	<ul style="list-style-type: none"> • shorter durations of repeated mentions of words, even when controlling for prosodic prominence • high frequency words exhibit more second mention reduction than low frequency words in plain speech, but not in clear speech 	
Baker and Bradlow (2007)	American English, Indian English, Korean	read paragraphs	<ul style="list-style-type: none"> • shortening of second mentions occurs even in languages with different prosodic systems 	
Baker et al. (2011)	native and non-native English	read paragraphs	<ul style="list-style-type: none"> • non-natives show tendencies similar to natives in the durational reduction of repeated mentions of content words • shorter word durations produced by native speakers, especially greater reduction of function words • higher within-speaker durational variance in native vs. non-natives 	non-native English produced by Chinese and Korean speakers

Koopmans-van Beinum and van Bergem (1989) investigated the effects of word repetition on vowel duration, F0 and vowel contrast, comparing spontaneous and read speech, as well as isolated words, obtained from one professional speaker. Vowel contrast was determined using the measure of the vowel distance to the centroid in a three-dimensional formant space (see Section 1.2.1.2 for details). The results showed no effect of word repetition on either of the inspected parameters in read speech. In spontaneous speech, however, both the F0 values and values of distance to the centroid were significantly higher in first mentions as compared to repeated mentions. The authors mention that the differences in F0 may be partly due to declination (in cases where both mentions of a word occurred within the same sentence), or may be related to amplitude differences. The study concludes that the spectral reduction of vowels is a better acoustic correlate of the reduction of repeated mentions of words than vowel duration.

A study by Shields and Balota (1991) used sentences with target words which were either (1) repetitions of an earlier mention of the same word, (2) associatively related to a previously mentioned word (e.g. *dog* and *cat*) or (3) not related to previous words within the same sentence. Speakers were instructed to produce the sentences from memory, as if they communicated the information to someone. The results showed that word durations were shortest in word repetitions, but associatively related words had also significantly shorter durations than words unrelated to a previous word in the sentence. In addition, peak amplitude (in dB) was found to be lower in word repetitions, when compared to the amplitude in the remaining two conditions.

The studies mentioned so far ascribe the higher degree of reduction of repeated mentions and their decrease in intelligibility when presented in isolation to the greater contextual redundancy of word repetitions. Hawkins and Warren (1994), on the other hand, assume that the differences in the degree of reduction, as well as intelligibility, are rather associated with the presence or absence of pitch accent. Such prosodic prominence tends to occur more often on words within an informational focus (i.e. first mentions of important content words). This study was based on material from spontaneous conversations, in which an experimenter guided the topic using pictures or photographs, and aimed to investigate the intelligibility of excised words and word fragments presented in isolation. Indeed, the results confirmed that the potential effect of the “new-given distinction” may be well explained by a large effect of pitch accent and its uneven distribution in the first and subsequent mentions of words (in particular, 93% of first mentions compared to only 46% of second mentions had pitch accent). According to the data in this study, word repetition did not have a significant effect on word intelligibility, when pitch accent was controlled for.

McAllister et al. (1994) paid more attention to the communicative context in which the repeated mentions occur. In this study, speakers were recorded while performing a problem-solving task (Tangram task) in monologue and in interaction with another speaker (dialogue). The results showed that speakers had a more conservative production strategy in monologue situations. In a dialogue condition, the word durations

Chapter 1

were overall shorter, and the durations of words' second mentions were significantly reduced compared to their first occurrences. In monologues, on the other hand, only a slight gradual decrease in word duration was found. The authors suggest that the more conservative production strategy in monologues may be due to missing verbal feedback.

A study by Fisher and Tokura (1995) showed that the reduction of repeated mentions of content words also occurs in infant-directed speech. The material included verbal descriptions of short puppet events that mothers gave to their 14-month old infants. Repeated mentions of target words were found to be shorter, quieter, lower-pitched and less variable in pitch. The authors assume that the acoustic difference between new and given words may serve to attract infants' attention to new words at the expense of background information.

Fowler et al. (1997) used spontaneous narratives of a movie content to investigate the shortening of repeated names and content words. They found the duration shortening of the repeated occurrences of words within but not between episodes. Follow-up experiments using read material showed that the shortening of repeated mentions is blocked particularly by metanarrative statements (i.e. explicit references to a scene change). The authors assume that this kind of episode boundaries causes a shift in the speaker's focus away from the story itself to the vehicle by which the events were conveyed to the speaker.

Gregory et al. (1999) investigated the influence of several factors related to word predictability on word duration as well as other reduction phenomena (i.e. frequency of realisation of word-final /t/ or /d/ as taps or their deletion). Apart from probabilistic factors reflecting the word frequency and collocational probabilities, the *number* of preceding mentions of a word within a discourse and a word's semantic relatedness to the preceding discourse were also found to contribute to word shortening. The other observed reduction phenomena turned out to be influenced by collocational probabilities, but no consistent influence of discourse-related predictability measures was found.

More detail regarding the influences on the shortening and decrease of intelligibility in repeated mentions of words in dialogues was provided by Bard et al. (2000). Their materials consisted of a series of Map Task dialogues (see Section 1.1.2 for details) where each instruction giver described a route through the same map in two consecutive sessions with different instruction followers. The results show word intelligibility decrease on the second mention, even when the word's first occurrence was directed to a different follower, or when the listener reported an inability to see the respective landmark. Moreover, the shortening and intelligibility decrease of the second mentions of words in a discourse took place regardless of which participant produced the first mention. The authors assume that word duration and intelligibility are simultaneously controlled by two types of mechanisms: fast automatic priming processes dependent on the speaker's knowledge, and slower optional processes that demand inference about the listener's knowledge.

Another attempt to clarify the role of prosody in the durational reduction of repeated mentions was made by Aylett and Turk (2004). They propose the Smooth Signal Redundancy Hypothesis, stating that speakers' articulation is affected by two opposing constraints: (1) producing robust communication and (2) efficiently expending articulatory effort. This can be expected to lead to an inverse relationship between language redundancy and duration, which enables to spread information more evenly across the speech signal. This is consistent with Lindblom's H&H theory (Lindblom 1990). In addition, the Smooth Signal Redundancy Hypothesis claims that prosodic prominence is the linguistic means to achieve the smooth signal redundancy. The study by Aylett and Turk (2004) investigated syllable durations (raw as well as normalised relative to the number of segments) in a corpus of Map Task dialogues (see Section 1.1.2 for details). The results showed a strong effect of several measures of language redundancy (including the number of previous mentions of a landmark), as well as prosodic factors, on syllable durations. Moreover, they found that the effects of prosodic prominence and redundancy factors are largely shared, although small unique effects of both prominence and redundancy were observed. The authors interpret this result as an

Chapter 1

indication that durational variation is for the most part controlled by prosodic structure, which “mirrors” language redundancy.

A study by Pluymaekers et al. (2005) inspected the seven most frequent Dutch words ending with an adjectival suffix *-lijk* to describe the influences of word repetition and predictability on the durations and number of realised segments in the word stem and suffix. The study excluded all items surrounded by pauses or disfluencies, and a number of other factors known to influence the degree of reduction (e.g. speech rate, presence of pitch accent, segmental context) were controlled for. The results indicated an effect of the number of previous mentions of a word on the duration of the word suffix, while the word stem duration and its number of segments were only influenced by the word predictability from neighbouring words.

Trón (2008) studied the influence of the number of previous mentions of a word within a dialogue and the time elapsed since a given word was mentioned previously within the dialogue on the durational reduction in pairs of consecutive mentions of the same word. The material was taken from the corpus of Map Task dialogues (see Section 1.1.2 for details). The results showed a greater durational reduction for shorter latencies¹³. In latencies over 60 seconds, the reduction asymptotically levelled out. The number of previous mentions of a word within a dialogue also influenced the magnitude of the reduction between the consecutive mentions. The largest reduction was thus found between the first and second mention of a word in a dialogue.

The question of whether the effects of word frequency and repeated mention interact with speaking style was addressed in a study by Baker and Bradlow (2009). The materials used in this study consisted of specially constructed texts produced in plain and clear speaking style. The results confirmed that words produced in clear speech have longer durations than those in plain speech, and high-frequency words and

¹³ Latency was defined as the time elapsed between the two consecutive mentions of the word in the comparison pair.

repeated mentions of words have generally shorter durations. The effect of the second mention reduction remained significant even in a data subset controlled for prosodic prominence. This indicates that second mention reduction is not a mere by-product of prosodic structure, reflecting the informational redundancy. In sum, all the investigated factors contributed to the resulting durations. In addition, a three-way interaction was found: high-frequency words exhibit more second mention reduction than low-frequency words in plain speech, but not in clear speech. The authors interpret this as a tendency to hypoarticulate when all the factors support it, with clear speech setting a limit on the amount of hypoarticulation allowed.

At this point, it needs to be mentioned that all of the studies discussed above have been carried out on English and Dutch materials produced by native speakers. To the best of our knowledge, very little is known about similar phenomena in other languages and in non-native productions. Baker and Bradlow (2007) investigated second mention reduction in read speech in Indian English¹⁴ and in Korean, compared with that in the productions of speakers of American English. The results confirmed that even in languages with different prosodic systems, the second mentions of words are significantly shorter than first mentions. In addition, it was shown that Indian English speakers were less likely to deaccent second mentions of words, which again contradicts the hypothesis that the durational reduction of second mentions is only due to the differences in prosodic structure. The only study that addresses the repeated mention effect in non-native production was carried out by Baker et al. (2011). Their material consisted of specially constructed paragraphs in English read by Chinese, Korean and native American English speakers. The study inspected the durations of content words and function words depending on their lexical frequency, and on whether they occurred previously within the discourse. In addition, foreign accent ratings were obtained for all

¹⁴ The Indian English speakers do not represent typical non-native English speakers due to the prominent role that English plays in the Indian educational system and society. All the speakers in this study reported that they were either native English speakers or learned English before the age of 6. They all reported to have one or more Indian languages as a mother tongue as well (Baker et al., 2011). The speakers could be possibly classified as bilingual speakers of this specific variety of English.

Chapter 1

non-native speakers and correlated with the investigated word-level durational features. The results showed that while non-natives exhibited similar tendencies as natives in the durational reduction of repeated mentions of content words, native speakers produced generally shorter word durations, showed especially greater reductions of function words, and had higher within-speaker durational variance than non-natives. Interestingly, no significant differences were found between the groups of Chinese and Korean speakers, possibly due to the generally higher between-speaker variance among non-native speakers. The foreign accent ratings correlated significantly with the within-speaker durational variance, the relative duration of function words and the overall word durations; non-native speakers with more native-like durations, greater within-speaker durational variance and greater reduction of function words were judged to have a less noticeable foreign accent.

In sum, the studies above showed that the effect of word repetition on shortening and other reduction phenomena, or on word intelligibility, is especially noticeable in spontaneous speech and communicative contexts. These effects were minimised in read speech as well as in monologues produced with a lack of listener's feedback. The fact that second mention reduction does not occur in word lists and across episode boundaries in narratives is a further indication of its communicative function. It could be speculated that second mention reduction is a listener-directed adaptive process. Some results, however, indicate that simpler¹⁵ priming processes are involved that interfere with the processes driven by the cognitively demanding inference about the listener's knowledge. The studies, however, also showed that second mention reduction is not due to simple articulatory priming, since it also occurs in words previously mentioned by the other conversation participant, but it does not occur in words preceded by homophones. Moreover, durational reduction also occurs to a certain degree in words semantically related to the preceding discourse. One point of controversy was the question of whether the durational reduction of repeated mentions is a direct effect independent of prosodic structure, or whether the durational differences between the

¹⁵ Depending on the speaker's knowledge

first and successive word mentions are controlled by means of the prosodic prominence structure. So far, several studies have shown that the second mention reduction effect remains significant even after controlling for prosodic prominence (Aylett and Turk, 2004; Pluymaekers et al., 2005; Baker and Bradlow, 2007; Baker and Bradlow, 2009), or that the reduction of successive mentions is gradual and sensitive to the number of previous mentions of the word within the discourse (Gregory et al., 1999; Aylett and Turk, 2004; Pluymaekers et al., 2005; Trón, 2008). At the same time, it is clear that the effect of prosodic structure on word durations is considerable and has to be taken into account in all the studies of other factors' effects on word durations. Finally, although most of the research on this topic has been carried out on English and Dutch materials produced by native speakers, initial evidence has already been brought showing that other languages, as well as non-native productions, may show the same tendencies in second mention reduction. Our work will supplement the existing research with new findings on similar phenomena in Czech, Norwegian and non-native English.

1.2.2 Relevant issues in non-native speech production

Up to now, research in non-native production has covered many areas, describing non-native segmental production as well as various aspects of prosody in non-native speech. Segmental studies have typically examined particular aspects of L2 segmental production (e.g. VOT durations in stop consonants, spectral and durational characteristics of vowels), trying to relate them to the structure of speakers' L1 segmental system. Studies on prosodic features in L2 production have described difficulties non-native speakers have with mastering various aspects of L2 prosody, such as stress, rhythm or intonation. Furthermore, some studies have attempted to investigate the contributions of prosodic and other features in non-native production to L2 speakers' perceived fluency, degree of foreign accent or the intelligibility of their L2 speech. To the best of our knowledge, however, few studies have focussed directly on the manifestations of phonetic reduction in non-native production. This section will give an overview of the studies more or less directly related to phonetic reduction, including investigations of global fluency-related phenomena, such as speech rate, and

Chapter 1

descriptions of the realisation of stress and rhythmical patterns in non-native production.

Nguyen and Ingram (2004) carried out a corpus-based study describing the incidence of a number of connected speech processes in non-native English produced by Vietnamese speakers, as compared to productions of native English speakers. They used a grammatical paraphrase task to elicit spontaneous speech from 12 native Australian English speakers and 11 Vietnamese speakers of English (with advanced level of proficiency). The results showed that several connected speech processes characteristic for spontaneous native English (i.e. consonant coalescence, liaison, certain types of consonant elision and vowel reduction) occurred significantly less frequently in English spoken by Vietnamese speakers. On the other hand, the results identified a number of processes occurring mainly in non-native speech as a result of the transfer of processes typical for the speakers' L1 (e.g. certain types of consonant deletion, vowel epenthesis in a consonant cluster or after a final consonant, initial implosive stops, lengthening of vowels in unstressed syllables, etc.).

A study by Bradlow et al. (2011) investigated several global features of speech timing comparing English, Mandarin Chinese and non-native English produced by Mandarin Chinese native speakers. The material included both scripted and spontaneous speech obtained from 11 English and 11 Mandarin Chinese speakers. The results showed that non-native English differs from native production (both English and Mandarin Chinese) especially by a lower speech rate (calculated as number of orthographic syllables per second) and a lower rate of syllable reduction (calculated as ratio of acoustic to orthographic syllables). At the same time, the speech rate and syllable reduction measures showed no significant differences between native English and Mandarin Chinese. Further, the analysis showed shorter speech chunks between silent pauses (calculated as number of words per silent pauses) in non-native as compared to native English. This feature may, however, be related to the speakers' L1 (Mandarin Chinese), where values similar to those of Mandarin-accented English were observed.

As we mentioned already in Section 1.2.1.3, a study by Baker et al. (2011) investigated the durational reduction of content words and function words depending on their frequency in the lexicon and whether they occurred previously within the discourse, comparing productions in native and non-native English. The study used read speech material from Chinese, Korean and native American English speakers. The results showed that while non-natives exhibited tendencies similar to natives in the durational reduction of repeated mentions of content words, native speakers generally produced shorter word durations, showed an especially greater reduction of function words, and had higher within-speaker durational variance than non-natives. Interestingly, no significant differences were found between the groups of Chinese and Korean speakers, possibly due to a generally higher between-speaker variance among non-native speakers.

A study by Granlund et al. (2012) addressed the issue of the adjustment of the articulatory effort in response to situational demands from the opposite perspective, by investigating the clear speech strategies in the production of Finnish-English late bilinguals. The material included spontaneous speech and a sentence reading task, the spontaneous clear speech being elicited naturally using a communication barrier (vocoding). The study focussed on the global enhancements of speech signal salience (i.e. mean energy between 1 and 3 kHz, F0 median and range, and speech rate), as well as on segmental modifications enhancing the phonological contrasts between categories (i.e. VOT of initial stop consonants, temporal and spectral characteristics for high front vowel contrasts). The results showed that speakers used largely similar global clear speech modifications in their L1 Finnish and their L2 English, and these were comparable to the global clear speech strategies used by native English speakers. With regard to segmental measures, the clear speech modifications of VOT in bilabial plosives were larger in English than in Finnish, presumably in order to increase the contrast between the English phonemes /p/ and /b/. In Finnish, on the other hand, the short-lag bilabial stop /p/ does not contrast in VOT with another category. Moreover, there was a tendency to a greater VOT contrast enhancement in more experienced speakers. As for the characteristics of high front vowel contrast, speakers were shown to

Chapter 1

use spectral and durational cues differently in the two languages, consistent with the differences in cue-weighting in the two languages. In clear speech, however, speakers used similar strategies for enhancing the spectral and temporal aspects of the vowels in both languages.

A number of studies have also described the phenomena related to phonetic reduction and global temporal parameters (such as speech rate and pausing patterns) in non-native production as part of speech fluency research. While some studies have described global fluency-related parameters comparing non-native vs. native productions (or productions of less vs. more experienced speakers), other investigations aimed at determining which objective measures contribute to the perceived fluency or degree of foreign accent (as evaluated by native speaker judges, etc.). The results of these studies may be particularly useful in suggesting the most important areas for improvement in L2 learners. It needs to be noted that although most of the global fluency-related parameters do not explicitly refer to reduction phenomena, reliable associations may be assumed. For example, the durational shortening of speech units (corresponding to a higher speech rate or articulation rate) has been shown to often imply articulatory attenuation resulting in various reduction phenomena (cf. Section 1.2.1). Moreover, an increased number of pauses in speech may hinder the occurrence of assimilations across word boundaries, or reduce the probability of segment deletions due to higher consonant cluster complexity.

A frequently investigated global temporal parameter is speech rate (or articulation rate), or an inversely related measure of matched sentence durations. A study by Riegenbach (1991) investigated several types of phenomena in productions of three very fluent and three very non-fluent non-native speakers of English (as judged by English instructors). In this study, a low speech rate was found to be rather typical for speakers judged as non-fluent, although one of the non-fluent speakers (“pseudo-fluent” speaker) was able to achieve a relatively high speech rate as well. Towell et al. (1996) investigated the development of several fluency measures in proficient French learners after a 6-month stay in a French-speaking country. This study showed an increase of both the speech

rate and articulation rate in the speakers' productions after a stay abroad. However, the values still did not reach the rate values achieved by the speakers in their L1 (English). Guion et al. (2000) compared sentence durations in the productions of native Italian and native Korean speakers that immigrated to Canada at different ages, as well as those of native English speakers. The speakers in this study repeated English sentences presented auditorily. The results showed that in both non-native groups, late bilinguals produced sentences of longer durations than the native English control group. In addition, a correlation between the sentence duration and the speakers' age of arrival was found. Cucchiarini et al. (2002) found that the articulation rate, as well as the scores of other fluency-related temporal measures, increased with the proficiency level in a reading task. In their samples of spontaneous speech, however, the opposite was found: the groups of beginner-level speakers had higher articulation rates than the intermediate speakers. This was explained by the different levels of difficulty in the tasks used for testing the beginner and intermediate speakers, resulting in a higher cognitive load for the intermediate speakers. MacKay and Flege (2004) compared the performances of early and late Italian–English bilinguals repeating matched English and Italian sentences following an aural model. The late bilinguals produced longer English than Italian sentences, whereas early bilinguals showed the opposite pattern. This result was interpreted as due to the late bilinguals' need for resources to suppress their Italian subsystem. Speech rate was also found to be a good correlate of fluency in Hungarian learners of English, as judged by three native English and three Hungarian teachers of English in a study by Kormos and Dénes (2004). The results also showed significantly lower speech rates for the less experienced speaker group. Trofimovich and Baker (2006) measured several suprasegmental parameters in the English production of three groups of Korean speakers with different levels of experience with English as L2. The results showed that the speech rates produced by all non-native speaker groups were significantly lower than in the control native English speaker group. Speech rate was also shown to (negatively) correlate with the speakers' age of arrival. Finally, Toivola et al. (2010) found significantly lower net articulation rates (i.e. syllables per second in a speech excluding repetitions and broken words) in the productions of Finnish learners as compared to those of native Finnish speakers, both in read speech and in task-elicited dialogues. In sum, lower speech rate or articulation rate (corresponding to longer word

Chapter 1

and sentence durations) seems to characterise speakers with less experience in a given language (e.g. non-natives vs. native speakers, late vs. early bilinguals). This may imply a lower degree of reduction in the speech of less experienced and less fluent speakers.

Apart from speech rate, certain pausing parameters have been consistently found to correlate with perceived fluency or speakers' L2 experience. For example, non-fluent speakers showed a higher pause frequency and a syntactically incorrect positioning of pauses in a study by Riegenbach (1991), and less experienced/less fluent speakers had longer mean durations of silent pauses (Towell et al., 1996; Kormos and Dénes, 2004). Trofimovich and Baker (2006) also found their least experienced speaker group to differ both in frequency and duration of pauses from the more experienced non-native speaker groups, as well as from the native control group. Other global temporal measures that have been shown to (positively) correlate with perceived fluency of speech include mean length of run¹⁶ (Towell et al., 1996; Cucchiariini et al., 2002; Kormos and Dénes, 2004), phonation-time ratio¹⁷ (Cucchiariini et al., 2002; Kormos and Dénes, 2004) and pace¹⁸ (Kormos and Dénes, 2004). While these measures may be particularly good correlates of fluency, they are obviously also strongly related to speech rate and silent pause frequency and duration.

A different approach to evaluating speech fluency was suggested by Hieke (1984). He describes the alteration processes occurring in running speech ("absorptions"), motivated by principles of ease of articulation. The mildest type of absorptions includes linking processes, such as "consonant attraction", i.e. the syllabic restructuring resulting from final consonants being attached to the following syllable if the syllable begins with a vowel. The study analysed the occurrence of this linking process in speech materials obtained by story retelling from native American English speakers and German students

¹⁶ Measured as the mean number of syllables or phonemes in a speech chunk delimited by silent pauses of certain duration

¹⁷ Percentage of time spent speaking out of the total time taken to produce the sample

¹⁸ Calculated as the number of stressed words per minute

of English. The results showed that while native speakers realised this kind of linking process in about 80% of the possible positions, non-native speakers only realised it in less than 54% of them. The frequency of occurrence of related processes can therefore be used as a reliable indicator of non-nativeness. Similarly, Bissiri and Volín (2010) compared the occurrence of the glottalisation of word-initial vowels in English spoken by natives and Czech speakers. Glottalisation in Czech productions occurred in a large majority of tokens (88% to 98%), with less dependence on the position (at phrase boundaries or at non-phrase boundaries), while English speakers glottalised only about 50% of the tokens at phrase boundaries, and even less at non-phrase boundaries (below 30%). These results seem to have an association with the more frequent occurrence of glottal stops before word-initial vowels in Czech. The presence of word-initial glottalisation phenomena may then be expected to hinder the occurrence of any linking phenomena or assimilations across word boundary.

Another area of L2 production research that relates to phonetic reduction is speech rhythm and lexical stress acquisition. The studies mentioned below deal with lexical stress positioning and realisation, as well as with the phonetic properties of unstressed syllables in non-native speech. In a study by Wenk (1985), trained native speakers evaluated the native-likeness of the reduced vowels in unstressed syllables in English produced by native French speakers with intermediate proficiency. While in a “sentence-final word echoing” task, the non-native speakers’ performance was practically native-like, the reduction of vowels in the (pre-tonic) unstressed syllables was only judged as native-like in 40% of cases in an imitative reading task, and over 60% in a guided retelling task. It needs to be noted that the judgements most likely depended on a wider range of cues than just the vowel quality reduction, including possibly the duration and pitch. Flege and Bohn (1989) investigated the placement and realisation of English stress in morphologically related words produced by native English and Spanish speakers. The results indicated the influences of word familiarity; the non-native realisations of high-frequency words were close to native-like. Apart from that, it was shown that while stress placement in the non-native productions was often correct, non-natives showed an insufficient degree of vowel reduction in

Chapter 1

unstressed syllables. A study by Nguyen and Ingram (2005) examined the production of English word stress by two groups of Vietnamese speakers with different amounts of experience. They found that both the F0 and intensity differences between stressed and unstressed syllables were successfully realised by non-native speakers. Native-like durational differences, however, were only correctly produced by the advanced speakers. The authors ascribe the beginners' failure to differentiate stressed and unstressed syllables in terms of duration to the fact that this phonetic feature is not active in their L1. Lee et al. (2006) studied the production of unstressed English vowels by early and late Korean-English and Japanese-English bilinguals. Apart from F0 and intensity, they measured vowel durations and formant values (F1 and F2). The results for vowel duration showed that both early and late Japanese bilinguals produced duration ratios¹⁹ similar to those of native English speakers, while both the early and late Korean bilinguals showed significantly less difference in vowel duration between their stressed and unstressed syllables, compared to natives. This difference between the speaker groups with different L1s may be due to the fact that Korean does not have any phonological length contrast, while Japanese does. Formant measurements showed that both groups of late bilinguals produced vowels more dispersed in the vowel space, with formant frequencies similar to full vowels with the same orthographic representation. The early Japanese group showed values similar to those of native English speakers with a slightly larger dispersion in the vowel space. In contrast to that, the early Korean bilinguals had the smallest dispersion for English unstressed vowel production which suggested that they used a native Korean vowel target /i/ instead of developing native-like formant values and variance. The fact that their unstressed vowel dispersion was smaller than that of the native English speakers is then consistent with the assumption that the English reduced vowels are subject to considerable coarticulatory effects.

Studies on speech rhythm acquisition may also provide insights into the non-native realisation of language-specific reduction phenomena. An impression of rhythm in

¹⁹ Ratio of the duration of the unstressed to the stressed vowel in a given word

general derives from the periodic recurrence of similar elements. Traditionally, languages were distinguished based on their speech rhythm into three classes: “syllable-timed”, “stress-timed” and “mora-timed” languages. The distinctions between these categories were assumed to be based on the isochrony²⁰ of some unit of speech. According to Abercrombie (1967), in languages with syllable-timed rhythm, syllables recur at similar intervals, while stress-timed languages demonstrate the isochrony of inter-stress intervals. In mora-timed languages, the successive morae²¹ are then assumed to have similar durations. This basis for the distinction of rhythm classes was, however, not supported by further research. Roach (1982) tested six languages previously used as examples of the two rhythm classes, and showed that (1) syllable duration variability is similar in both classes, and (2) inter-stress intervals are not more regular in languages representing the stress-timed rhythm category. Moreover, Dauer (1983) confirmed that stresses do not recur more regularly in English than in the four other observed languages, including Spanish and Italian. Instead, her data suggested that a tendency for stress isochrony was rather a universal property of the temporal organisation of languages. Furthermore, she observed that certain phonetic and phonotactic regularities of syllable structure tend to co-occur in languages from the same rhythm classes. Thus, the difference in the perceived speech rhythm seems to result from a number of language-related properties, including the distribution of different syllable structures, the possibility of vowel reduction in unstressed syllables and the phonetic realisation of stress. More recently, research on speech rhythm has proposed a number of acoustic metrics based on the durations of consonantal and vocalic intervals in the speech signal. Some of these measures reflect the overall signal properties, such as the average proportion of vocalic intervals within the speech signal (%V) and the average standard deviations of vocalic or consonantal intervals (ΔV , ΔC) (Ramus et al., 1999), while others attempt to capture the sequential nature of rhythmical contrasts by measuring the variability of consecutive consonantal and vocalic intervals (i.e. Pairwise Variability

²⁰ Temporal similarity

²¹ Mora is a minimal timing unit larger than a single segment and typically smaller than a syllable. It allows us to classify syllables in “light” (containing one mora) and “heavy” (containing two morae).

Chapter 1

Indices²²; see Grabe and Low, 2002). The following paragraph presents some of the relevant studies of L2 speech rhythm.

Bond and Fokes (1985) studied patterns of word compression due to the addition of syllable suffixes in non-native English. The material included four English words and their forms derived using mono- and disyllabic suffixes, read by native speakers of Japanese, Malaysian and Thai. The results showed that while the production of one of the Thai speakers matched quite well the native English patterns of word compression, the other Thai speaker as well as all the native speakers of Japanese and Malaysian showed an insufficient awareness of typical English word compression patterns. Although these speakers did shorten some of the words with added suffixes, the compression was not proportional to the number of added syllables, as in native English. The native-like production of one of the Thai speakers seems to be well explained by the rhythmical similarity of Thai with English which may facilitate the acquisition of rhythmical patterns. Gut (2003) investigated non-native speech rhythm in German using material obtained by reading and retelling a story. The participants included native speakers of Chinese, Italian and Polish, as well as a native German control group. A comparison of the percentage of vowel reductions and deletions in one type of German inflection suffix (/Cen/, /Cem/) showed a much lower frequency of vowel deletion by non-natives. This was especially noticeable in the productions of native Chinese speakers, possibly due to the different prosodic organisation of Chinese. Unlike the native Germans, the non-natives also produced the suffix with a full vowel in some cases. In addition to that, another measure of reduction, the mean ratio of duration of consecutive full and reduced syllables, was calculated. In read speech, the native productions showed significantly higher full/reduced ratios than all the non-native groups, while in story retellings, only the Chinese speaker group had values significantly lower than the natives. A study by Volin (2005) focussed on vowel durations in seven selected polysyllabic English words produced (within meaningful

²² Average of the differences in duration of successive consonantal/vocalic intervals

texts) by natives²³, and Czech speakers judged as having a strong non-native accent. The results showed that the Czech speakers had a considerably lower durational reduction coefficient²⁴ than the natives. It turned out that in most cases, the Czech speakers' stressed vowel durations were shorter than their mean unstressed vowel durations. A detailed analysis of selected items then revealed additional factors that seem to influence the durational patterns in the Czech speakers' productions (e.g. phonological length of some of the vowels in the Czech language equivalents of the studied words). In their previously mentioned study of global suprasegmental parameters in the production of non-native English speakers of different proficiency levels, Trofimovich and Baker (2006) also included the measure "stress-timing", defined as a ratio of mean unstressed syllable duration to mean stressed syllable duration. This measure describing the speakers' acquisition of L2 specific speech rhythm was shown to correlate with their amount of L2 experience (measured as the speakers' length of residence in an English-speaking country). A study by van Dommelen (2007) inspected the temporal patterns in Norwegian as a second language compared with those in native Norwegian. The study investigated read speech in Norwegian produced by speakers from six different language backgrounds, as well as by native Norwegians. To describe speech rhythm, the following seven metrics were used: syllable duration mean and standard deviation, correlation coefficient and regression slope for the relation between the syllable durations in each studied utterance and the rank numbers of the syllables according to the native Norwegian reference²⁵, speech rate mean and standard deviation (in phonemes per second) and normalised PVI for syllables. To determine whether the defined rhythm metrics capture the L1-specific deviations of speech rhythm, a discriminant analysis using these measures was carried out. The overall correct classification rate amounted to nearly 93%, indicating that non-native Norwegian differs rhythmically in language-specific ways from native Norwegian. Further, it was shown that only some of the measures contributed

²³ Professional newsreaders

²⁴ Duration of the stressed vowel relative to mean duration of the unstressed vowels within the word

²⁵ Within each studied utterance, the occurring syllables were ordered based on their mean durations in the Norwegian reference group, thus assigning a "rank number" to each of the syllables in an utterance.

Chapter 1

significantly to this classification, in particular the mean syllable duration, standard deviation for speech rate, correlation coefficient of durations of different syllable types relative to the Norwegian reference, and mean speech rate. Clearly, most of these measures refer to the rate of speech production, which possibly does not represent an L1-specific factor. The correlation coefficient measure, on the other hand, seems to be a useful rhythm measure reflecting the systematic deviation of the durations of various syllable types from native Norwegian reference. White and Mattys (2007) used a number of rhythm metrics (interval measures as well as pairwise variability indices) to compare the rhythmical properties of native vs. non-native English, Dutch and Spanish read speech. The inspection of rhythm metrics for non-native productions, where the speakers' L1 and L2 were from different rhythm classes, revealed interesting patterns of percentage of vocalic intervals within the signal (%V), and also of the speech-rate normalised variability of vocalic intervals (VarcoV²⁶). Both the native Spanish speakers' English productions and the native English speakers' Spanish productions showed intermediate VarcoV values between the native productions of the two languages from different rhythm classes. On the other hand, only the Spanish speakers' English productions showed intermediate values of %V, while the English speakers' Spanish productions seemed to have overshot the target (native Spanish) value. Their very high proportion of vocalic intervals in speech may be explained by the greater effects of accentual and phrase-final lengthening in combination with a higher frequency of open syllables in Spanish. In contrast to that, non-native productions from speakers whose L1 and L2 were both stress-timed languages (Dutch and English) did not differ significantly in any of the interval variability measures. However, small differences in %V values were found between the two languages as spoken by natives. Interestingly, even in non-native productions the %V values were closer to the speakers' L1 values, suggesting that the accommodation between rhythmically similar languages may not be undertaken by most of the speakers. A corpus study by Gut (2007) compared native and non-native productions in German and English, using different types of speech material ranging from word lists to free speech in an interview situation.

²⁶ Variation coefficient of vocalic interval duration

The non-native speaker groups for both studied languages included a large number of speakers (~50) representing a variety of native languages. The results showed significantly lower syllable ratios²⁷ in the non-native vs. native productions (both in German and in English). In addition, the productions of non-native speakers of English contained a significantly lower percentage of reduced or deleted syllables as compared to those of the native English speakers. Grenon and White (2008) used several rhythm metrics (%V, VarcoV, rPVI_C²⁸) to compare the productions in native and non-native English and Japanese. Interestingly, although Japanese speakers' non-native English appeared comparable to native English in all chosen metrics, a closer analysis revealed that the Japanese speakers' ratios of vowel duration in consecutive stressed and unstressed syllables were much lower than those of the native English speakers, indicating an insufficient durational reduction of unstressed syllables. On the other hand, the English speakers' Japanese production showed significantly higher rPVI_C values, most probably as a result of the English speakers' aspiration of voiceless Japanese consonants.

The studies mentioned in this section have found initial evidence that certain linking and assimilation processes found commonly in native spontaneous productions occur less frequently in non-native productions (Nguyen and Ingram, 2004; Hieke, 1984). Similarly, a considerably lower ratio of syllable deletion was observed in non-native vs. native speech by Bradlow et al. (2011). On the other hand, non-native productions in a given language may display a more frequent occurrence of processes that normally do not occur in the language spoken by native speakers, but are transferred from speakers' L1 (Nguyen and Ingram, 2004; Bissiri and Volín, 2010). Furthermore, several frequently studied fluency parameters of non-native speech, such as speech rate and articulation rate, may be considered indirect indicators of the degree of reduction in speech. Generally, lower speech rate or articulation rate (corresponding to longer word and sentence durations) seems to characterise speakers with less experience in a given

²⁷ Mean ratio of duration of consecutive full and reduced syllables

²⁸ Pairwise variability index for successive consonantal intervals, not normalised for speech rate

Chapter 1

language (e.g. non-natives vs. native speakers, less experienced vs. more experienced L2 speakers, late vs. early bilinguals). In contrast to that, in certain previously identified reduction-inducing contexts, non-native productions may show tendencies similar to native speech. Baker et al. (2011) showed that repeated mentions of content words within the discourse tend to get shortened by natives as well as non-native speakers. The within-speaker durational variance is, however, smaller in non-natives than in natives, indicating that the non-natives' durational reduction fails to reach the degree achieved by natives. Less experienced speakers' productions also seem to display an increased frequency or duration of silent pauses, or their inappropriate positioning (not on syntactic constituent boundaries), as well as a higher frequency of other phenomena disrupting speech fluency (e.g. glottalisation, lack of linking). A likely consequence of the higher frequency of fluency-disrupting phenomena is a decreased occurrence of word-linking, assimilations across word boundary and other related phenomena motivated by the reduction of articulatory effort in connected speech.

Several studies on the non-native realisation of stressed vs. unstressed syllable contrast in English came to similar conclusions indicating an insufficient vowel quality reduction in the unstressed syllables by learners or late bilinguals (Wenk, 1985; Flege and Bohn, 1989; Lee et al., 2006). On the other hand, early Korean bilinguals in the study by Lee et al. (2006) produced vowel formant values in English unstressed syllables with much less variance than native English speakers, apparently as a result of using a Korean vowel target instead of acquiring native-like reduced vowel values and variance. The durational reduction of unstressed syllables was also found to be inappropriate in less advanced learners (Nguyen and Ingram, 2005), and in productions by native Korean learners of English (Lee et al., 2006). As for the acquisition of L2 rhythmical patterns, a number of studies have confirmed smaller durational contrasts between the stressed and unstressed syllables in non-native vs. native production, or in the production of less experienced speakers (i.e. speakers with a shorter length of residence) (Gut, 2003; Volín, 2005; Trofimovich and Baker, 2006; Gut, 2007; Grenon and White, 2008). The above mentioned results refer to the rhythm acquisition of stress-timed languages (German and English). Further, studies of some concrete phenomena

showed that non-native productions of rhythmical patterns deviate from those of natives, possibly depending on the similarity of rhythmical properties of the speakers' L1 and the target language (Bond and Fokes, 1985; Gut, 2003). With regard to more abstract rhythm metrics, productions of L2 learners, whose L1 and L2 were from different rhythm classes, showed values of the rhythm metric VarcoV intermediate between the native values of the two languages from the different rhythm-classes. The VarcoV values of the productions of L2 learners, whose L1 and L2 were from the same rhythm class, on the other hand, did not differ from native productions. Interestingly, in the case where values of the rhythm metric %V only differed slightly between the speakers' L1 and L2 (the two languages were from the same rhythm class), non-native speakers did not accommodate to this difference, and retained the same %V values as in their L1 (White and Mattys, 2007). Finally, two of the studies mention the effect of speaking style on some aspects of non-native speakers' production. The results show a greater deviation of non-native production from the native rhythmical patterns in read speech, as compared to the free speech obtained from story retellings (Wenk, 1985; Gut, 2003). Although this tendency may seem rather surprising, assuming that read speech production poses less cognitive demands on non-native speakers (by sparing them the effort to formulate the message, find appropriate lexical items, etc.), a possible explanation is the freedom of choice of the structures familiar to the speaker in free speech production. In reading tasks, on the other hand, non-native speakers may be faced with more complex linguistic structures whose production may be particularly unnatural for less experienced speakers, resulting in a less successful implementation of L2 rhythmical patterns.

The results of the above summarised studies on the reduction-related phenomena in non-native production seem to converge on several points. Firstly, it seems that L2 speakers have more difficulty with acquiring L2 processes that do not exist in their native language, or accommodating their productions to match patterns that differ considerably from those occurring in their native language. Generally, more experienced learners are more likely to master those aspects of the L2. Apart from that, phenomena transferred from the speakers' L1 may occur in non-native productions.

Chapter 1

Some evidence also seems to suggest that the successful acquisition of a certain aspect of L2 production may be hindered by the fact that the learner's L1 is rather similar (but not completely) in that aspect. A fine accommodation to L2 may then be avoided by retaining production patterns characteristic of the speaker's native language. These conclusions seem to be in line with the well-known hypotheses of Flege's Speech Learning Model (1995). Although the Speech Learning Model (SLM) is primarily concerned with L2 sound acquisition, it seems reasonable to relate some of its hypotheses to other phenomena on a general level. In short, this model assumes that the phonetic systems used for language learning remain adaptive throughout the speakers' lifespan, and can be used for L2 learning. The model describes the development of phonetic categories and the interaction of L1 and L2 phonetic categories in the shared phonological space. According to the model, a new category can be established for an L2 sound as long as the cross-language phonetic difference between the L2 sound and the closest L1 sound can be discerned by the L2 learner. The likelihood that the cross-language phonetic difference is discerned depends on the degree of perceived dissimilarity, and on the age when the speaker has initiated L2 learning. Finally, L2 sounds similar to L1 sounds will be perceptually assimilated to L1 categories, resulting in the inhibition of new category formation. The previously summarised evidence from studies on the non-native production of reduction-related phenomena and rhythmical patterns seems to suggest that factors such as the similarity of structures in L1 and L2, and the amount of the L2 learner's experience with L2 are relevant in ways predicted by the SLM, even for areas beyond L2 sound acquisition. It needs to be taken into account, however, that there are a number of additional factors involved in L2 acquisition (e.g. familiarity with other languages, formal training in L2, education level, motivational factors, etc.) that may obscure possible systematic effects (cf. van Dommelen, 2007: 138).

1.3 Selected characteristics of the studied languages

In this section we will give a brief overview of some phonetic and phonological characteristics of the three languages that will be the object of the study. The aim is to

briefly describe aspects that may be relevant to the investigated phenomena associated with phonetic reduction. First, this overview will describe the rhythm types of the three languages and related aspects of their phonological and phonetic characteristics, including the phonological length distinction, syllable structure and properties of lexical stress (Section 1.3.1). Section 1.3.2 will mention some connected speech processes reported to occur in the studied languages. In Section 1.3.3, the phonetic characteristics associated with the phonological voicing contrast and voicing assimilation phenomena in the three languages will be described in more detail.

1.3.1 Rhythm type and related phonological properties

Regarding the rhythm type, both English and Norwegian are classified as stress-timed languages, while the rhythm classification of Czech has been a subject of some discussion. While traditional descriptions classified Czech as a syllable-timed language (Palková, 1997: 285), a more recent study by Dankovičová and Dellwo (2007) pointed out that depending on the rhythm measure used, Czech may show characteristics of stress-timing (e.g. a high variability of the duration of consonantal intervals), and should be therefore considered a mixed type.

An aspect known to be relevant to rhythmical properties of a language is the syllable structure. Languages traditionally classified as stress-timed typically show more complex syllable structures. In English, particularly stressed syllables are found to have a complex structure, and the overall ratio of open syllables is only 44% (Dauer, 1983). Syllables may contain up to three consonants in syllable onset and up to four consonants in coda (e.g. Jensen, 1993: 65-70). Likewise, a Norwegian syllable may have complex onsets and codas. Syllable onsets may contain up to three consonants, while codas (in morphologically complex words) may contain up to five consonants (Moen and Kristoffersen, 2006). Complex syllables may also occur in Czech, both in stressed and unstressed positions. Syllable onsets in Czech may contain up to four consonants, while syllable codas contain a maximum of three consonants (e.g. Palková, 1997: 271). However, the proportion of open syllables in Czech is relatively high (73%, cf. Kučera 1968: 50; cited in Palková, 1997: 273).

Chapter 1

Another aspect potentially relevant to language rhythm, as well as to possible manifestations of phonetic reduction, is the presence of the phonological quantity contrast of vowels. Although English monophthongs are traditionally divided in short (lax) and long (tense) vowels, and a part of the vowel system can thus be organised in pairs of vowels with corresponding quality, the acoustic cues distinguishing between the short and long member of each pair are not limited to duration. The main cue for this distinction is the spectral quality of the vowels. Vowel duration is further used to signal other distinctions (e.g. voicing of preceding consonant). In Norwegian, a vowel quantity distinction occurs in syllables with word stress. In addition, consonants in stressed syllables have a complementary distribution of duration, so that a short vowel is followed by a long consonant or a consonant group, and a long vowel by a short consonant. Unstressed syllables, on the other hand, can only contain short vowels. In Czech, the phonological contrast between short and long vowels occurs both in stressed and unstressed syllables. The short vs. long member of each vowel pair do not differ in quality with the exception of the pair /ɪ/ and /i:/ (e.g. Palková, 1997: 171). This suggests that in Czech, vowel duration is the only cue to a phonological contrast in a wide range of contexts.

As for the realisation of lexical stress, syllables with primary stress in English tend to be longer and louder compared to other syllables, and often marked by a pitch excursion. Moreover, vowels in unstressed syllables undergo qualitative reduction. Lexical stress in English is not bound to a fixed position within the word, and in some cases the stress placement may distinguish words that are segmentally identical. In Norwegian, syllables with lexical stress have several qualities that unstressed syllables do not possess. Stressed syllables may contain a long segment in the rhyme (i.e. a long vowel or a long consonant in the coda), while unstressed syllables only contain short segments. In addition, syllables with primary stress are the domain of tonal accents²⁹ (e.g. Holm,

²⁹ Tonal accents in Norwegian are two distinctive F0 patterns. An accent contour stretches over a stressed syllable and at least one following unstressed syllable. The tonal accents have a distinctive function in a part of the lexicon (e.g. Holm, 2008).

2008). Vowels in unstressed syllables may also undergo qualitative reduction. In Czech, lexical stress has a delimitative function, being placed on the first syllable of a word or a preceding (monosyllabic) preposition. Although the description of Czech stress is an area of continuing research, describing the overall F0 course throughout the stress group seems to be more promising than attempts to identify any prosodic properties bound to the stressed syllables (Palková and Volín, 2003). Neither the intensity nor duration (which is used mainly to signal the phonological length in vowels) has a decisive role in signalling stressed syllables, although they may contribute to the perceptual impression of stress to a certain degree. Moreover, the quality of vowels occurring in unstressed syllables in Czech is not reduced (e.g. Palková, 1997: 278-279).

1.3.2 Connected speech processes

In English, apart from vowel reduction in lexically unstressed syllables, a number of other casual speech reduction phenomena have been described that occur in unstressed syllables and further increase the contrast between stressed and unstressed syllables (Shockey, 2003: 22-32). Schwa absorption describes cases where a schwa disappears, while another sound in its vicinity takes on its syllabic property. The resulting element may be a sonorant, a fricative or even a voiceless vowel. Schwa suppression is a phenomenon describing schwa assimilation by a neighbouring vowel resulting in a loss of its syllabicity. Further, reduction of closure for obstruents occurs particularly often in syllables immediately following a stressed syllable. This type of lenition includes a loss of closure in stops and reduced approximation in fricatives. Tapping of alveolar stops or clusters is a process resulting in very short consonants, produced with a lower control over the articulation movements. Although this phenomenon is typical especially for American, Australian and Irish English varieties, it occurs occasionally in SSBE as well. Moreover, due to the more demanding aerodynamic-articulatory conditions of the production of voiced obstruents, phonetic voicing in obstruents is avoided where possible. In unstressed syllables, phonologically voiced stops are rarely phonetically voiced, and also fricatives are mostly at least partially devoiced. This occurs frequently (but not only) in a context of following voiceless consonants. Conversely, phonologically voiceless segments are sometimes produced as voiced (lenis), especially

Chapter 1

in intervocalic positions. This is more likely to happen to continuant consonants, including those that became continuants by the lenition of stops. A number of processes have also been described that minimise the sequences of consonants or vowels and restructure the sequence of sounds towards a more regular alternation of consonants and vowels, which is assumed to be more natural as well as articulatorily less demanding (Shockey, 2003: 32-44). In order to avoid the adjacency of two vowels, a linking [r] may be inserted between two consecutive vowels in careful speech. In casual speech, on the other hand, processes like [v/0] alternation³⁰ in the word *of* and /l/-vocalisation³¹ may be used to preserve a CV-type syllable structure. Other modifications of syllable codas include the realisation of /t/ as a glottal stop, and the elision of /d/ between two consonants (most often after /n/ or /l/). A very common process is nasal relocation. This can occur when a nasal is followed by a homorganic obstruent, particularly a voiceless stop. Due to a modification of the timing of the velum lowering gesture, the nasal consonant is only reflected in the nasalisation of the preceding vowel. /h/-dropping is especially prevalent in certain accents of English, but in most English accents, it may occur in short unstressed words following a consonant (especially a fricative). In weak forms, this process is part of standard pronunciation. Finally, several processes have been reported that can be classified as articulatory assimilations. /ð/-reduction may affect words starting with /ð/ that are preceded by an alveolar consonant. /ð/ is then assimilated to the preceding alveolar consonant, usually in the articulation manner or voicing (Shockey, 2003: 43-44). Gimson (1980: 290-291) also mentions regressive assimilation of the articulation place of word-final alveolar consonants, which occurs frequently in colloquial speech, and coalescence³² of alveolar obstruents with /j/.

³⁰ Alternation between the pronunciation of the preposition *of* as [əv] and [ə], where the latter is typically produced when followed by a consonant

³¹ Loss of tongue contact in word-final velarised /l/ before consonants

³² Coalescence (or fusion) is a process in which two segments occurring in a sequence combine into a single segment, usually exhibiting some characteristics of both of the original segments. It may be regarded as a type of assimilation.

To the best of our knowledge, no systematic description of reduction processes in Norwegian has been provided. However, Broch (1935) lists a number of cases of drastic reductions or “contractions” occurring in rapid speech in East Norwegian. The examples serve to illustrate the rhythmical principle that guides the restructuring of sound-matter in disyllabic stress-groups composed of a stressed (heavy) syllable followed by an unstressed (light) one. A large part of the discussion focusses on frequently occurring constructions formed by an auxiliary verb and negative particle, adverb or pronoun, which form fixed combinations (“fixed clichés”) that are particularly prone to reduction as a result of the rhythmical principle. According to the observations presented in this paper, syncope occurs frequently in unstressed syllables in words containing three or more syllables, either manifesting as deletion of the whole syllable or only the vowel, which then leads to creation of more complex, often otherwise unusual consonant clusters. Vowels of two syllables may also be fused into a single syllable without forming a recognised diphthong (synizesis). On the other hand, disyllabic words that form a stress group on their own do not undergo such reduction. Here, vowels in unstressed syllables may be reduced to schwa or syllabic consonants, but the number of syllables is not affected. Apparently, factors such as speech rate, speaking style, speaker’s status and education, but also the context and the possibility of misunderstanding, play an important role in determining the degree of reduction.

The following contact assimilation phenomena and other connected speech phenomena known to occur in Czech are described based on Palková (1997: 144-147, 323-338). Apart from regressive voicing assimilation which will be discussed in more detail in Section 1.3.3, assimilations of articulation place and manner occur in Czech. While manner assimilation is rather rare, (regressive) assimilation of the place of articulation occurs regularly and in some cases it is even obligatory. The occurrence of the articulation place assimilations across word boundary is, however, considered a feature of substandard production. Other assimilation (accommodation) phenomena occurring in casual speech in Czech include the nasalisation of vowels adjacent to a nasal consonant and intervocalic consonant lenition (causing change of stops into fricatives or approximants, producing a voiceless sound as voiced, or a complete disappearance of a

Chapter 1

sound). However, the occurrence of these phenomena is typically classified as a feature of substandard pronunciation or a regional dialect. Other described connected speech phenomena include consonant elisions resulting in consonant cluster simplification. Here, Czech orthoepic rules define which particular elisions are allowed in the spoken standard to avoid meaning confusion or a drastic reduction of intelligibility. In substandard production, an even wider variety of elisions may occur. With regard to vowels, the shortening of phonologically long vowels in word endings was observed to occur in casual substandard production. In some of the regional dialects, shortening may occur even in long vowels in word stems. Although in Czech, vowel reduction in unstressed syllables does not occur, vowel quality may be reduced as a result of fast or sloppy pronunciation. This tends to happen more frequently in vowels located between consonants with a similar place of articulation, typically in longer words. Vowel quality reduction may cause a significant reduction of intelligibility and it takes place only in substandard production. Apart from the connected speech processes discussed above, it seems relevant to mention certain other phenomena which occur in Czech, and whose function is to facilitate the segmentation of the speech stream into words. These include primarily the presence of a glottal stop (or glottalisation) before a word-initial vowel, or word medially after a prefix. Glottal stops occur automatically after a pause. Moreover, the pronunciation norm requires a glottal stop use between a non-syllabic preposition and a word-initial vowel, and in a range of other contexts it is recommended for the sake of intelligibility. Its absence is typical for some regional dialects and for careless speech. Glottal stops in Czech also trigger regressive voicing assimilation. In some contexts (preceding /o/), a word-initial prothetic /v/ may occur in place of a glottal stop in substandard pronunciation.

1.3.3 Voicing contrast and voicing assimilation

In English, the distinction between phonologically voiced and voiceless consonants is signalled by a number of phonetic characteristics. Depending on the type and position of the consonant, they include the duration of the preceding vowel, the duration of the consonant itself or some of its articulatory phases, and the presence of vocal fold vibration (cf. Morland, 2010: 8-9). The duration of the consonant, as well as the

duration of the preceding vowel, is used to signal (phonological) voicing in Norwegian, too (e.g. Fintoft, 1961; van Dommelen and Ringen, 2007). However, vocal fold vibration (phonetic voicing) seems to be a very reliable indicator distinguishing between voiced (lenis) and voiceless (fortis) consonants in Norwegian as well (cf. a study of intervocalic stops by van Dommelen and Ringen, 2007). It seems that in spite of the presence of additional phonetic cues to phonological voicing, the role of phonetic voicing as a cue to the phonological voicing contrast in Norwegian remains relatively important. The relatively lower importance of durational cues may be due to the fact that they are primarily used to signal the phonological length of vowels. In Czech, the phonological voicing contrast is primarily signalled by the presence or absence of phonetic voicing, which is perceptually crucial (Machač and Skarnitzl, 2007). Although moderate durational differences between voiced and voiceless consonants exist as well, their perceptual relevance has not been confirmed so far (Skarnitzl, 2011: 104-105).

According to traditional descriptions, regressive voicing assimilation in English occurs only to a limited degree. In particular, regressive assimilation may occur across word boundaries, but it is only triggered by tense (voiceless) obstruents (e.g. Roach, 1983: 106; Jansen, 2007). However, evidence has been brought indicating that certain forms of regressive voicing assimilation in English occur in a broader range of contexts, including assimilations triggered by the lax obstruents /d/ and /z/. It needs to be mentioned that these assimilatory effects influence the amount of phonetic voicing, but do not cause a complete neutralisation of the phonological voicing contrast (Jansen, 2007). Further, progressive voicing assimilation in English occurs in cases where an -s suffix (for a noun plural or a verb in third person singular) is attached to a word (e.g. *cats* will be pronounced as /kæts/ and *dogs* as /dɒgz/; Roach, 1983: 107). Moreover, lenis obstruents in initial and final positions often have little or no phonetic voicing, although the phonological voicing contrast is not neutralised (Roach, 1983: 30-31, 38). In Norwegian, with a few exceptions, post-vocalic obstruent clusters consist of sounds matching in voicing. Kristoffersen (2000: 74-79) documents the tendency to the regressive (neutralising) devoicing of stem-final obstruents when adding an adjectival suffix -t, and less regularly when adding various suffixes starting with /s/. In addition,

Chapter 1

both the progressive and regressive devoicing of non-nasal sonorants adjacent to voiceless obstruents takes place quite regularly (Kristoffersen, 2000: 76, 79). In Norwegian, voicing contrast is maintained even in word-final positions (e.g. Husby and Kløve, 2001: 62). It has to be mentioned, however, that the role of phonetic voicing in the signalling of the voicing contrast in word-final position has not been investigated (cf. Morland, 2010: 10). In Czech, all obstruents are organised in pairs of corresponding³³ voiced and voiceless sound³⁴, and in obstruent clusters they undergo obligatory regressive assimilation of voicing within words as well as across word boundaries (e.g. Palková, 1997: 213, 328-329). An exception is the labiodental fricative /v/ which does undergo assimilation, but does not trigger it with a regularity compared to other obstruents (cf. Skarnitzl, 2011: 124-135). Another noteworthy exception is the fricative trill, which undergoes progressive assimilation after voiceless obstruents (Palková, 1997: 330). All obstruents are also subject to word-final (prepausal) devoicing (e.g. Palková, 1997: 329). Although traditionally both the assimilation of voicing and word-final devoicing are described as complete neutralisations of voicing contrast, more recent evidence suggests that this distinction is only partly neutralised (Podlipský and Chládková, 2007).

1.4 Aims and hypotheses

1.4.1 Aims

In the last decades, the research on phonetic reduction has brought a wealth of results describing the effects of various factors on the occurrence of reduction phenomena, as well as explaining the mechanisms of particular types of reduction (see Section 1.2.1). However, the majority of such research has been carried out using various types of native speech materials, while only very few studies included non-native productions.

³³ Having the same manner and similar place of articulation, with the exception of the pair formed by the voiced phoneme /ř/ and voiceless /x/

³⁴ Some of the sounds only have an allophonic status.

The present thesis work aims to complement the existing research by focussing on reduction phenomena in non-native speech. The two studies constituting this thesis attempt to describe selected durational and spectral parameters across domains typically associated with varying degrees of phonetic reduction, using materials in the non-native English produced by Czech and Norwegian speakers as well as native productions.

The goal of the first study is to describe the effect of speaking style on realisations of the English function words *in*, *of* and *to* in the productions of native speakers and two groups of non-native speakers (Czech and Norwegian speakers). The investigated speaking styles include read speech and spontaneous dialogues elicited using a conversational task. The parameters under observation include both temporal and spectral properties. Moreover, several aspects of context are taken into consideration, including the articulation rate and type of neighbouring segment.

The second study included in this thesis addresses the reduction of repeated mentions of content words in the course of a dialogue comparing native productions (in Czech, English and Norwegian) and productions in non-native English spoken by native speakers of Czech and Norwegian using task-elicited dialogues. The investigation includes both temporal parameters, and the spectral characteristics of the vowel in the stressed syllable. In addition, an attempt is made to control for factors related to the discourse status and prosodic structure.

1.4.2 Hypotheses

According to Lindblom's H&H theory (Lindblom, 1990), phonetic reduction may be viewed as a result of the general tendency to minimise articulatory effort, and thus certain forms of reduction may be expected to occur in all the studied languages, as well as in non-native productions. On the other hand, previous research has shown that phonetic reduction occurs to a smaller degree in non-native production, compared to that of native speakers. Studies of fluency parameters, which may be considered indirect indicators of the degree of reduction in speech (see Section 1.2.2 for details), have shown that lower articulation rates and a more frequent occurrence of fluency-

Chapter 1

disrupting phenomena seem to characterise speakers with less experience in a given language (e.g. non-natives vs. native speakers, less experienced vs. more experienced L2 learners, late vs. early bilinguals). Therefore, it may be expected that the degree of reduction will be lower in non-native as compared to native production, and the amount of the speakers' L2 experience will be an important factor influencing the degree of reduction in the speakers' L2 production. A number of studies in the area of L2 acquisition have also confirmed that less experienced speakers have difficulty with acquiring L2 processes that do not exist in their native language, or accommodating the productions to match patterns that differ considerably from those in their native language, while phenomena similar to those existing in the speakers' L1 are more readily acquired. At the same time, phenomena transferred from the speakers' L1 may be found to occur in non-native productions, even though they do not occur in native productions in the target language. In light of these findings, we may expect that the degree to which particular reduction-related phenomena are employed by a non-native speaker depends on whether and to what degree similar phenomena occur in the speaker's L1.

Based on the literature reviewed in the previous sections, we present the following general hypotheses:

- (H0) Non-native speakers will display tendencies to phonetic reduction similar to those displayed by natives
- (HA1) Reduction in non-native productions occurs in a smaller degree than in native production, and its amount depends on the amount of the speaker's experience with the given L2
- (HA2) Particular reduction patterns and other aspects of phonetic realisations in non-native productions may deviate from the native-like patterns in ways that can be traced back to characteristics of the speaker's L1

The present investigation will address some of the issues relating to these general hypotheses, although the scope of this research may not allow a systematic and detailed

exploration of all facets of the hypotheses. The conclusions of the research may therefore not provide clear and in-depth findings in relation to all the relevant aspects of the above mentioned hypotheses.

1.5 Thesis outline

Chapter 2 provides details about the speech materials used throughout this thesis and information about the speakers. Chapter 3 describes a study of the realisations of three English function words (*in*, *of* and *to*) produced by native and non-native speakers. To that aim, the study compares the productions in two speaking styles: read and spontaneous speech, and inspects a number of acoustic measures relating both to the temporal organisation and to the spectral properties of the observed function words. The influence of several context-related variables is also investigated. The chapter contains a detailed description of the methods used in the study, followed by a presentation of the results for each function word. In conclusion, a summary of the main findings and their discussion with relevant literature are provided. Chapter 4 aims to investigate the phonetic reduction of repeated mentions of content words in native and non-native productions, focussing on durational, rhythmical and spectral aspects of reduction. In addition, several presumably influential factors related to the discourse status of the observed words, prosodic structure and articulation rate are taken into consideration. The chapter first introduces the study methods, and proceeds with the description of the results of the statistical analyses. The most important findings are summarised and discussed at the end of the chapter. In Chapter 5, the main findings of the studies described in Chapters 3 and 4 are summarised, and general conclusions are presented (in relation to the hypotheses formulated above). Lastly, a number of methodological issues identified in the course of the investigation are mentioned, and some directions for future research are suggested.

Appendix A presents samples of pictorial material relating to the task used for speech material collection. Appendices B and C contain additional statistical information relating to the variables investigated in Chapter 3 and Chapter 4, respectively. Appendix

Chapter 1

D presents a table listing all the lexical items occurring in the study of the reduction of repeated mentions of content words in Czech, English and Norwegian (Chapter 4) based on the number of a word's syllables. Appendices E and F contain the detailed results of the statistical analyses carried out to model the effects of selected control factors, as well as the main experimental factors, on word duration and spectral contrast in the observed content words.

2 Speech materials

In this section, the different types of material used for the studies carried out in the framework of this thesis will be described. An important part of the material used in this investigation is spontaneous conversational speech. The selection of method to obtain spontaneous speech recordings is a complicated issue, and various approaches to this challenge were exemplified in Section 1.1.2. The next section (Section 2.1) discusses some of the advantages and disadvantages of using conversational tasks to elicit spontaneous speech, and explains the reasons for choosing this method for collecting spontaneous material for this thesis. Details about the speakers participating in the recordings can be found in Section 2.2. The recording of the conversational material used throughout this thesis is described in Section 2.3. In addition, one of the studies compares the parameters in spontaneous speech with read speech. For this purpose, read speech was collected from the same speakers that produced the spontaneous material. The recordings of read speech will be described in Section 2.4.

2.1 Task-elicited spontaneous speech

As was shown in examples in Section 1.1.2, there are several obvious advantages of using tasks for eliciting spontaneous speech. First of all, assigning tasks to speakers offers an opportunity for eliciting speech in the speakers' non-native language. The instruction to perform a task in a non-native language does not influence the naturalness of speakers' speech behaviour during the task to a large extent. On the contrary, asking two speakers that share the same native language to speak freely in a non-native language could be perceived as a very difficult and unnatural activity. In addition, the assigned task may captivate the speakers' attention and motivate them to achieve a good result. This helps to distract their attention from the fact that they are being recorded, and thus reduces the risk of unnatural or reticent speech behaviour due to the speakers' uneasiness in the recording situation. Another advantage of conversational tasks is that they offer a certain degree of control over the content and structure of the conversation, which may be important for quantitative research. It is also possible to modify a task by,

Chapter 2

for example, asking the speakers to use particular words or expressions (e.g. using special landmark names in the Map Task as described in White et al., 2010). The session may include an initial training phase to guarantee that the speakers are familiarised with the expressions and are able to produce them fluently in the course of a conversation (without having to read the landmark names from the map). On the other hand, there are also some problems resulting from using tasks to elicit spontaneous conversations. One of the disadvantages is a potential asymmetry of the conversation resulting from the different roles of the speakers in the task. Some of the tasks may also cause the speakers to use certain stereotypic utterance patterns, reducing the natural variability of conversational speech.

2.2 *The speakers*

The recordings used throughout this thesis were obtained from ten native Czech speakers, ten native Norwegian speakers and ten native English speakers. All the speakers were non-professionals, i.e. did not have any special speech performance (enunciation) training. Most speakers were university students, between 19 and 35 years of age. Most of the native English speakers spoke Standard Southern British English. Three speakers were from other regions within Great Britain, but did not have a strong regional accent. In the Czech and Norwegian recordings, various dialects of Czech and Norwegian are represented. Details about the individual speakers, their dialects, and the onset of English exposure for the non-native speakers of English (i.e. in most cases the age when they started learning English at school), and other information related to the speakers' use of English, can be found in Table 2.1. The speakers were paid a small amount in compensation for their participation in the recordings.

Table 2.1: Speaker data, including native language (L1), pair and speaker identifiers, age (years), gender (m=male, f=female), region(s) of speaker dialect, the onset of speaker's exposure to English (years of age), and other information relevant to the speaker's English experience.

L1	Pair #	Speaker	Age	Gender	Dialect group	L2 experience start	Other L2 related information
Norwegian	1	EA	28	m	Trøndelag	10	
	1	AM	26	m	West Norway	bilingual	bilingual: mother is American, but they lived in Norway
	2	AH	22	m	Trøndelag / East Norway	5	holidays in English-speaking countries
	2	MBG	23	f	Trøndelag	7	holidays in English-speaking countries
	3	NFH	19	f	East Norway	10	2 years at an English school
	3	MBE	19	m	East Norway	3	lived in the USA from 3 to 7,5; 2 years at an English school
	4	JOO	22	m	West Norway	8	
	4	EV	23	m	East Norway	8	
	5	IET	25	f	East Norway / North Norway	10	
	5	MSE	25	f	East Norway	10	
Czech	6	VH	20	m	South-West Bohemia	10	
	6	JT	20	m	South-West Bohemia	6	
	7	Aka	21	f	Central Bohemia	11	1 year in the USA
	7	MA	20	m	Central Bohemia	10	
	8	KV	21	f	Central Bohemia	11	two 14-day courses in England/Ireland
	8	MS	20	f	Central Bohemia	11	frequent short (week) stays in the UK
	9	TE	21	m	Central Bohemia	3	working in an English-speaking company (1 year)
	9	MJ	35	m	Central Bohemia	15	working in English-speaking companies (5 years)
	10	JP	21	f	East Moravia / Central Bohemia	9	
	10	Ako	20	f	Central Bohemia	13	
English	11	JE	25	f	SSBE		
	11	GMH	29	f	SSBE		
	12	SG	26	f	SSBE		
	12	AW	27	f	SSBE		
	13	PD	21	m	SSBE		
	13	VS	32	f	SSBE		
	14	TJ	32	m	Midlands		
	14	KP	26	m	SSBE		
	15	RA	27	f	South Wales		
15	IM	23	f	Yorkshire			

2.2.1 Selection of non-native speakers of English

Since the Czech and Norwegian speakers were supposed to produce read and spontaneous speech in English, an evaluation of their proficiency in English would be relevant to guarantee a homogeneous population. Testing the proficiency of the non-native speakers prior to the recording was, however, not possible. This was compensated by a deliberate selection of subjects. In Norway, the well-established system of English instruction and a high exposure to the English language (e.g. most movies in English screened in Norway are not dubbed) result in an overall high competence in English in the young population. The speakers were therefore selected from university students, regardless of the subject of their study, which guaranteed sufficient proficiency. In the Czech Republic, however, such a proficiency standard cannot be generally expected, and we had to select speakers from more carefully chosen groups, namely university students of English, and employees in a company using English as the official work language. As a result of these selection criteria, a sufficient level of proficiency was guaranteed for all the speakers. In addition, the speakers' confidence in being able to perform the conversational task in English³⁵ was implicit in their decision to participate in the recordings. Consequently, an assessment of selected fluency-related temporal parameters in the spontaneous English production of the speakers was carried out to obtain rough information about one of the aspects of the speakers' L2 proficiency (see Section 2.2.2).

From the data presented in Table 2.1 we can summarise, that the Czech speakers were on average slightly younger (mean age 21.9 years) than the Norwegian speakers (23.2 years). Combined with the Czech speakers' later start of English instruction (9.9 years vs. 7.1 years for Norwegians), the time span between the start of English instruction and the time of recording was considerably shorter (12 years vs. 16.1 years, for Czech and Norwegian speakers, respectively). These differences may possibly explain some of the differences between the two non-native speaker groups observed in the following

³⁵ Brief information about the conversational task used for speech elicitation and the requirements on the speakers' English proficiency were provided to the speakers prior to scheduling the recording sessions.

chapters. However, due to the limited number of speakers in this study, and the multitude of confounded factors influencing the L2 performance (see for example Piske et al., 2001; Flege and MacKay, 2010), it may not be possible to draw any general conclusions relating to these L2-instruction related factors.

Apart from the L2 proficiency-related factors, we should note that in our material, the speech produced by non-native speakers was also *directed* to non-native interlocutors. In addition, the two speakers in each non-native speaker pair had the same native language background. We may speculate that the fact that the speakers shared the native phonological system might have contributed to a decrease of the speakers' efforts to achieve the L2 phoneme targets. This would result in a certain exaggeration of some characteristic features of their non-native accent. However, issues of this type can hardly be avoided. It is well known that speakers tend to adapt their productions to the needs of their audience (cf. Uther et al., 2007; Scarborough et al., 2007), and therefore, it is rather difficult to obtain a completely "neutral" speech.

2.2.2 Assessment of non-native speakers' fluency

As we mentioned in the previous section, in spite of the careful selection of non-native speakers of English, we could observe noticeable differences in some of the observed factors related to their experience with English as an L2, not only between the individual speakers but also between the groups of Czech and Norwegian speakers. As a result of that, the level of proficiency in the two speaker groups may be expected to differ. However, we are aware that due to the relatively small number of the speakers in our sample and the impossibility to consider all factors that have a potential influence on a speakers' L2 performance, such a prediction may not be too reliable. To obtain a more accurate estimate, we carried out an analysis of the actual data related to speaker proficiency.

Clearly, the level of language proficiency manifests itself in many aspects of language use, including the appropriateness of the use of grammatical structures, the precision in lexical choice, the degree of native-likeness in the realisation of L2 sounds and

Chapter 2

suprasegmental patterns, as well as the speech fluency. In addition, the individual speakers' performance in each of these aspects may vary considerably. Therefore, any metric or test of language proficiency can only indirectly refer to the abstract level of language proficiency. Although typically, language proficiency evaluation involves written tests, and thus provides information mainly about the speakers' knowledge of grammar or the wealth of their vocabulary, we decided to focus on selected fluency-related temporal parameters, which are more closely related to spoken performance. The relevance of this approach has been confirmed by previous research which found evidence that certain global temporal parameters of speech, such as articulation rate, pause frequency and duration, typically correlate with the amount of the speakers' experience in a non-native language, and are thus well-suited for evaluating a speakers' overall proficiency (see Section 1.2.2 for a summary).

The assessment of fluency-related temporal parameters in the spontaneous English production of Czech and Norwegian speakers was carried out using 3 to 9 monologue speech stretches³⁶ per speaker. The duration of these stretches amounted to a total of 57 to 90 seconds per speaker. In this material, we annotated all silent pauses with a duration of at least 100 ms as well as filled pauses (hesitation sounds) not completely integrated into a word³⁷, and calculated the number of syllables uttered within the speech intervals (including repairs and false starts). Using these basic data, we calculated the speakers' *articulation rate* as the number of syllables per second within clean speech intervals (excluding filled pauses), *phonation ratio* as the percentage of vocal activity (i.e. speech or filled pause) within the total duration of monologue intervals, and *mean length of run (s)* as the mean duration of speech chunk delimited by silent pauses. In addition, *mean length of run (syll)* reports the mean number of syllables

³⁶ Monologue speech stretches were defined as intervals within the spontaneous dialogue consisting of the coherent speech activity of one of the speakers, only interrupted by backchannels or very short responses from the other speaker.

³⁷ For example, an extremely lengthened vowel at the end of a word would not be labelled as a filled pause, unless a noticeable vowel quality change is present.

a speaker produces as one chunk between silent pauses. Table 2.2 lists the values for each speaker as well as the means for Czech and Norwegian speakers.

Table 2.2: Selected global temporal parameters in the spontaneous English production of the non-native speakers

L1	Pair #	Speaker	Articulation rate (syll/s)	Phonation ratio (%)	Mean length of run (s)	Mean length of run (syll)
Norwegian	1	EA	3.6	59.0	0.9	2.9
	1	AM	5.4	77.8	1.3	6.6
	2	AH	5.0	73.9	1.9	8.0
	2	MBG	4.3	86.2	2.0	7.6
	3	NFH	4.6	73.5	1.3	5.9
	3	MBE	4.6	78.9	2.0	8.2
	4	JOO	4.0	70.0	1.3	4.2
	4	EV	3.2	72.5	1.5	4.2
	5	IET	4.4	73.5	1.8	7.4
	5	MSE	4.4	80.9	1.7	7.0
		Mean	4.4	74.6	1.6	6.2
Czech	6	VH	4.3	79.6	1.3	5.1
	6	JT	4.7	81.7	1.6	7.2
	7	AKa	3.7	77.7	1.4	4.5
	7	MA	4.7	71.8	0.8	3.9
	8	KV	3.3	84.1	2.5	7.8
	8	MS	3.9	76.5	1.4	5.3
	9	TE	4.5	82.0	1.9	8.0
	9	MJ	4.9	70.5	1.2	5.1
	10	JP	3.9	76.5	1.6	5.0
	10	AKo	3.7	78.4	1.5	4.7
		Mean	4.2	77.9	1.5	5.7

We can observe that the differences between the two speaker groups are rather slight. The mean articulation rate of Czech speakers is 4.2 syllables per second compared to 4.4 syllables per second in the Norwegian speaker group. On the other hand, the mean phonation ratio is slightly higher in the group of Czech speakers (77.9%) than in the Norwegian speaker group (74.6%), indicating a smaller proportion of silent pauses in the overall speech performance of Czech speakers. The values of the mean length of run in seconds show very little difference between the two groups (1.5 s vs. 1.6 s for Czech

Chapter 2

and Norwegian speakers, respectively), indicating only a slightly higher pause frequency in the Czech speakers' productions. The measure of the mean length of run in syllables combines both the information about the pause frequency and articulation rate within the speech intervals, and therefore, it shows the most noticeable difference between the Czech and Norwegian speaker groups, namely 5.7 syllables per chunk for the Czech speakers, compared to 6.2 syllables per chunk for the Norwegians. Figure 2.1 shows a boxplot of the mean length of run in syllables for the Czech and Norwegian speakers. The boxes in the plot contain the inter-quartile range, and the whiskers extend to the furthest datapoint that lies within 1.5 times the inter-quartile range from the box on each side. It is apparent that even though the median values of the two groups vary considerably, the variation between the speakers within each group is substantially higher. Moreover, statistical tests confirmed that neither of the observed fluency-related parameters differed significantly between the two speaker groups.

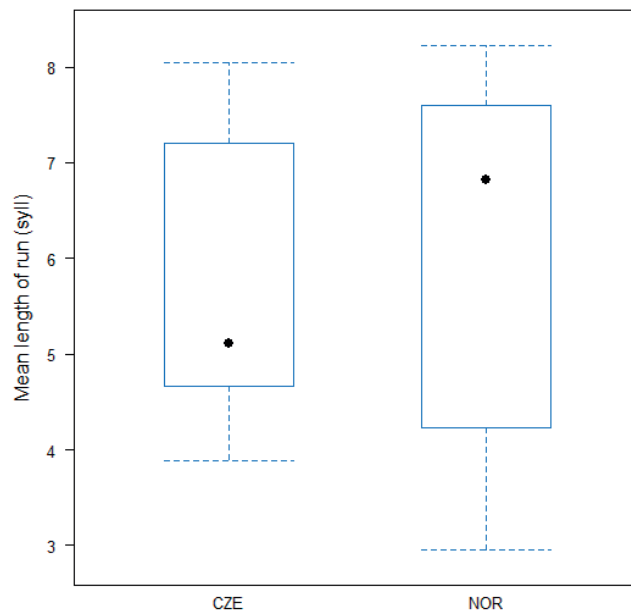


Figure 2.1: Mean length of run (in syllables) for the two groups of non-native speakers of English with different L1 backgrounds (CZE=Czech, NOR=Norwegian).

Lastly, we examined the correlations of the above mentioned fluency measures with two variables relating to the speakers' experience with English as a second language: their onset of English exposure (i.e. in most cases the age when they started learning English at school) and the time span between the start of English instruction and the time of recording (i.e. number of years they were learning or able to use English). It should be pointed out that both of these variables were rather roughly related to the speakers' actual experience with English, since we did not control for the extent and quality of their English instruction, nor the extent of English use later on. In addition, as we mentioned above, many other factors have been found to have an effect on L2 performance, rendering the relationship between these L2 experience-related factors and performance even less transparent.

There was a significant correlation of the articulation rate with the speakers' onset of English exposure ($r = -0.464$; $N = 20$; $p = 0.039$), as well as with the time span between the start of English instruction and the time of recording ($r = 0.533$; $N = 20$; $p = 0.016$). Apart from that, a significant correlation was found between the speakers' onset of English exposure and their mean length of run in syllables ($r = -0.513$; $N = 20$; $p = 0.021$). The remaining correlations of fluency measures with variables relating to the speakers' experience with English as an L2, however, did not reach significance. Still, the relationships of the articulation rate and mean length of run in syllables with the relevant variables relating to the speakers' L2 experience seem to be quite reliable, considering the significance of the correlations in a sample with only 20 subjects. To illustrate this, the relationship between the articulation rate and the time span between the start of English instruction and the time of recording for the two groups of non-native speakers of English with different L1 backgrounds is shown in Figure 2.2.

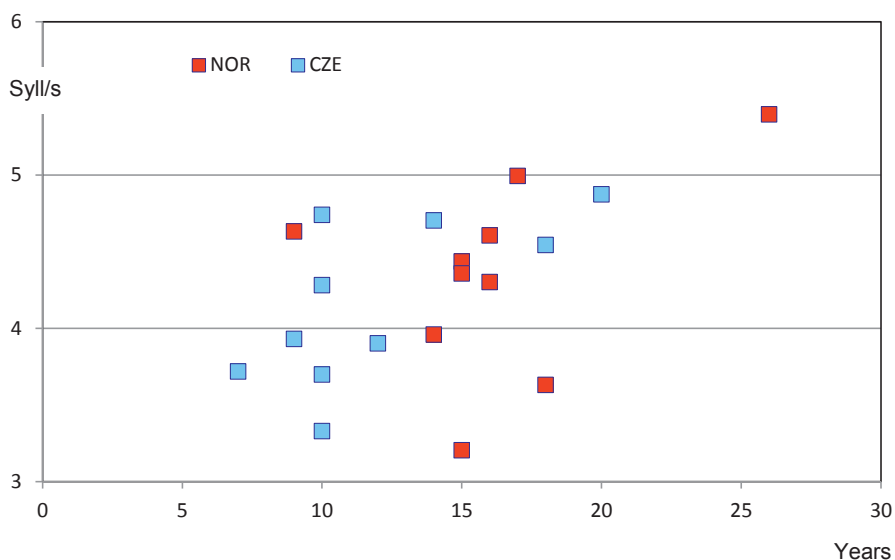


Figure 2.2: Relationship between articulation rate (syll/s) and the time span between the start of English instruction and the time of recording (years) for the two groups of non-native speakers of English with different L1 backgrounds (CZE=Czech, NOR=Norwegian).

2.3 Spontaneous speech recordings

The spontaneous speech material used in this thesis consisted of recordings in Czech, Norwegian, and native and non-native English. The recordings in non-native English were produced by Czech and Norwegian speakers. The recordings in Czech, Norwegian and non-native English are a part of the Kachna corpus (Spilková et al., 2010). The recordings were obtained from the group of 30 speakers described in Section 2.2. For the conversational speech recordings, speakers participated in pairs, as indicated in Table 2.1 on page 66. In most cases, the speaker pairs were formed by either classmates or colleagues.

2.3.1 Elicitation task and instructions

The conversational speech recordings used in the thesis are based on a “picture replication task”. In this task, one speaker receives a detailed cartoon illustration

depicting a humorous situation (adapted from Butschkow, 2002). This speaker is instructed to describe the picture to the other speaker, whose task is to replicate it as accurately as possible on a sheet of paper using a pencil. In order to encourage the active participation of the drawing speaker in the dialogue, an accompanying task was added: the sheet for drawing contained five detail sections cut out from the original picture and the drawing speaker was instructed to identify their content and determine their location within the picture. Neither of the speakers could see the other's picture. The speakers were asked to use approximately 30 - 40 minutes for the task. Appendix A contains one of the illustrations used for the picture replication task (reproduced with the permission of the publisher Baumhaus Verlag), an empty sheet for drawing and two examples of the resulting drawings.

In Czech, the standard language differs from the colloquial or dialectal varieties. In formal situations (e.g. in lectures, university examinations and the media) the standard variety is considered appropriate, but for the Kachna corpus recordings, speakers were explicitly encouraged to use their dialectal or colloquial varieties of Czech, i.e. to speak together the way they normally would in their everyday interactions. In Norway, on the contrary, the use of dialects is generally encouraged, and dialects are considered appropriate in all social situations. Due to this status of Norwegian dialects, no instructions of this kind were necessary for the Norwegian speakers. The English speakers were not instructed as to the particular use of their dialects. To avoid undesirable dialect variability in the native English material (which was supposed to serve as a reference), a more careful selection of speakers was carried out, excluding speakers with a strong non-standard accent.

As a result of the task, the recorded conversations were quite lively. In most of the recordings, the describing speakers were more active in the dialogue, while the drawing speakers mainly asked questions for clarification. The speakers seemed to be motivated by the task objective, and in many cases also amused by the comical drawings. This contributed to the relaxed atmosphere and naturalness of the conversations. In some cases, the speakers became so immersed in the task activity that they significantly

Chapter 2

exceeded the recommended duration of the recording. A summary of the recording durations is given in Table 2.3.

Table 2.3: Recording durations (minutes)

L1	Pair #	Language	Duration (min)	Language	Duration (min)
Norwegian	1	English	53	Norwegian	34
	2		48		38
	3		49		21
	4		57		34
	5		73		38
	Total		280		164
Czech	6	English	41	Czech	29
	7		38		27
	8		31		30
	9		35		25
	10		32		41
	Total		177		153
English	11	English	27		
	12		33		
	13		36		
	14		41		
	15		65		
	Total		201		

The actual amount of usable speech material per speaker, however, made up only a part of the total recording duration and depended on the speakers' role in the task ("describer" or "drawer"), their eloquence and the total duration of the dialogue. According to a rough estimate, the durations of speech for the majority of speakers were between 6 and 20 minutes. A few (2) of the speakers provided as little as 3.5 minutes of speech material.

2.3.2 Recording sessions

For the native speakers of English, the structure of the recording session was simple. The speakers were instructed about the objectives of the task and encouraged to choose

their roles, and then they carried out the task. The describing speaker is listed as first within the speaker pair in Table 2.1 on page 66.

The non-native English speakers were recorded both in English and in their native language. Therefore, the recording sessions were considerably longer. The structure of the recording session was the same for all the non-native English speaker pairs. After the speakers were instructed in the task, the session started by recording the speakers performing the task in English (their second language) and proceeded until the speakers were satisfied with their achievement. The describing speaker in this dialogue is listed as first within the speaker pair in Table 2.1. After the first recording, a short refreshment break followed where the speakers could amuse themselves by comparing the model picture and the resulting drawing. Subsequently, they carried out the same task in their native language. For this second recording, a new picture was provided, and the speakers exchanged roles, so that the describer role for this recording was taken by the speaker listed as second within the pair in Table 2.1. It needs to be mentioned that due to the speakers' fixed roles in the task within each part of the recording, the amount and richness of collected speech material inevitably differed between the speakers within a pair. However, as mentioned in Section 2.3.1, the drawing speaker obtained an additional task designed to reduce the possible imbalance in the dialogue.

2.3.3 Audio recording and processing

The dialogues were recorded in a sound-treated studio at the Department of Language and Communication Studies, NTNU, Trondheim, for the Norwegian speakers, and at the Institute of Phonetics, Charles University in Prague, for the Czech speakers. One pair of the native English speakers was also recorded in the studio in Trondheim, while the rest of the English speakers were recorded in a sound-attenuated booth at the Department of Experimental Psychology, University of Bristol. Table 2.4 lists the technical parameters for the recordings made at the three locations.

The dialogues were recorded in stereo (one channel per speaker), with a sampling rate of 44.1 kHz and 16-bit quantisation. The recordings in Trondheim and Prague were

Chapter 2

obtained using two boom-mounted microphones. This choice can guarantee a high sound quality of the recordings, using microphones with excellent technical parameters. Due to the activities involved in the replication task, the speakers were not expected to move significantly relative to the microphone, and therefore, the use of headsets was not considered necessary. In Bristol, however, head-mounted directional microphones were used. This solution helps to maintain sufficient channel separation in dialogues.

Table 2.4: Recording and audio processing specifications

	Trondheim	Prague	Bristol
Audio format	wav	wav	wav
Microphone	MILAB LSR-1000	AKG C 4500 B-BC	headset Shure WH20
Pre-processing	high-pass filter (50 Hz)		
Sound card	Creative SB Live	Sound Blaster Audigy 4	?
Sampling rate	44.1 kHz	44.1 kHz	44.1 kHz
Quantisation	16 bit	16 bit	16 bit
Audio-processing software	Adobe Audition version 2	Sound Audio Studio 8.0	Cool Edit

The recording settings were consistent within the three locations. However, the dimensions of the recording studios were dissimilar. The size of the studio in Trondheim allowed for the speakers to sit several meters apart, facing in different directions, each at a different table. The studio in Prague is much smaller, so it was necessary for speakers to sit next to each other at the same table, partially visually separated by a styrofoam board (obscuring one another's pictures from view during the task). Similarly, the booth in Bristol was very small, and the speakers had to sit opposite each other at one table, separated by a felt-coated board in the middle. The positions of the speakers in different studios are depicted in Figure 2.3.

As a result of the different studio layouts and different microphone types, the channel separation differs among the recordings. Recordings of Norwegian and English speakers (made in Trondheim and in Bristol) have channels very well separated, while those of the Czech speakers have a strong cross-channel overlap. Due to this, the portions of overlapping speech in the Czech recordings were disqualified for use in detailed

phonetic analyses. Moreover, an obvious consequence of the recording studio layout and the activities involved in the replication task was a reduced possibility for visual face-to-face interaction between the speakers. The naturalness of the resulting entirely acoustic communication is still adequate, and can be compared to telephone conversations. In addition, the absence of visual contact makes speakers fully rely on acoustic cues, which makes the recordings well suited for the investigation of acoustic cues in speech interaction.

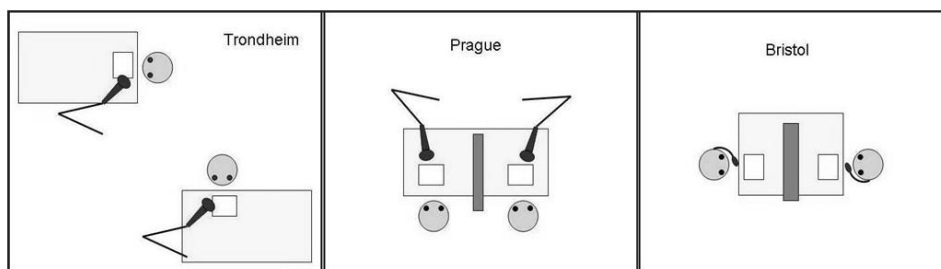


Figure 2.3: The positions of speakers during recording in the studios in Trondheim (left), Prague (middle) and Bristol (right). The sketches do not illustrate the dimensions of the studios (they were not the same, see text).

2.4 Read speech recordings

The recordings of read speech were obtained from all the speakers described in Section 2.2. All the English speakers and eight of the Norwegian speakers were recorded on the same day of the spontaneous speech recording, just before carrying out the picture replication task. For eight of the Czech speakers, who were the students of English at Charles University, the read speech recording took place several months earlier in connection with their seminar participation. These recordings were kindly provided by the Institute of Phonetics, Charles University in Prague. The read speech of both the two remaining Czech speakers and the two remaining Norwegian speakers was recorded approximately 18 months after the recording of their spontaneous dialogues.

Chapter 2

The reading task consisted of reading one page of a BBC news transcript. The speakers were allowed to read the text before the start of the recording and were instructed not to worry excessively about the pronunciation of unusual foreign names that were likely to affect the reading fluency. The speakers only read the text once. For most of the speakers, the read speech recording durations were between 3 and 5 minutes.

3 Realisations of English function words in read and spontaneous speech

3.1 Introduction

As presented in the literature overview in Section 1.2.1.2, speaking style research so far has addressed various topics, including durational and spectral properties of speech sounds, the frequency of some types of reduction phenomena and various prosodic characteristics of speech productions in different speaking styles. Most of the studies have focussed on differences between spontaneous speech and read speech, while other defined speaking styles (e.g. clear speech) have been investigated less often. The results of previous research have shown relatively consistent tendencies with regard to the spectral properties of speech sounds realised in different speaking styles, namely a higher degree of reduction (e.g. vowel centralisation, reduced cepstral differences) and more variability in more spontaneous speaking styles. Parameters relating to prosody and temporal measures, on the other hand, showed often diverging results both across the studies and between speakers within a study. While some of the discrepancies are undoubtedly due to the differences in material types used in the different studies, speaker-specific strategies to achieve a given speaking style are assumed to play an important role as well. Lastly, it should be pointed out that there has been very little research investigating speaking style differences in non-native production. Cucchiarini et al. (2002) measured a number of fluency-related parameters in read and spontaneous speech in Dutch produced by non-native speakers with different proficiencies, showing longer pause durations and a higher frequency of pauses and disfluencies in spontaneous speech. However, the spontaneous speech obtained from the different proficiency groups resulted from different tasks (presumably presenting different cognitive demands to the speakers), and therefore, the results cannot be directly compared between the groups. Moreover, we are not aware of any research comparing native and non-native productions in read and spontaneous speaking style.

Chapter 3

The present study aims to address this topic by inspecting the realisations of three English function words (*in*, *of* and *to*) using native and non-native English material (non-native being produced by Czech and Norwegian speakers) in read and spontaneous speaking styles. The study addresses the temporal organisation as well as the spectral characteristics of the three words. Moreover, local articulation rate and some aspects of segmental context are taken into consideration.

3.2 Method

3.2.1 Speech material

The material used in this study consisted of read speech and spontaneous task-elicited dialogues in English, produced by native British English speakers and two groups of non-native speakers of English (Czech, Norwegian). Both types of material, as well as details about the speakers, were described in Chapter 2.

3.2.2 Selected items

The lexical items chosen for analysis were three English function words: *in*, *of* and *to*. They were chosen because they belonged to the most frequent lexical items in the material. For both the read and spontaneous material, we aimed to select realisations of these words fluently and naturally integrated in surrounding speech; therefore, we excluded all cases where a pause, hesitation or another type of disfluency was present in close proximity to the observed word. There were, however, a number of speakers whose (L2) production was characterised by careful, hesitant pronunciation. In those cases, we even included items that showed some marks of hesitation proportional to the speaker's standard. We also avoided utterances marked by overly precise pronunciation due to, e.g. introducing a new or unexpected lexical element. Attention was also paid to the context and syntactic status of the observed words, where we avoided, e.g. use of words in the initial or final position in an utterance (i.e. not surrounded by context), clause-final use of prepositions (so-called *stranded prepositions*) and strongly lexicalised phrases where a disproportional reduction could be expected. In addition,

when selecting the tokens of the function word *to*, we avoided most of the cases where the word was preceded by an alveolar plosive (/t/ or /d/). In such situations it is very common that the two sounds are realised as one geminate, and therefore, it is not possible to accurately determine the portion representing the first segment of the word *to*. Naturally, we also excluded cases where a disturbing background noise occurred that influenced the quality of the speech signal (e.g. sounds of pencil scratching on the paper, or noises caused by speakers touching their microphones).

Five tokens of each word per speaker and speaking style were selected (incidentally less than five for a few speakers with a limited number of suitable items). Whenever there were more than five suitable tokens available, the first five were used. In total, we investigated 868 tokens in this study. Each token was stored in a separate sound file, including approximately 1 second of context on each side.

3.2.3 Acoustic analysis and additional observations

To prepare the selected items for automatic data extraction, the items were segmented using Praat (Boersma and Weenink, 2009). In order to achieve consistency in the segmentations, a number of criteria were taken into account, following the recommendations in Machač and Skarnitzl (2009a). In particular, the full formant structure in vowels was exploited as a basic cue for the segmentation of sequences containing a vowel and a fricative, plosive or nasal. Furthermore, the presence of friction in the spectrogram, and waveform amplitude and shape were considered, as well as the auditory impression. In cases with gradual transitions between segments, the boundary was placed in the middle of the transition phase (see Machač and Skarnitzl, 2009a: 35-36). It needs to be mentioned that vowel portions with creaky phonation or other glottalisation phenomena, occurring quite frequently on vowel-vowel transitions, were often segmented as part of the observed vowel. In these cases, the quality in the creaky portion was considered, and an auditory evaluation of the items was of particular importance. Excluding creaky and glottalised portions of vowels may result in disproportionately shorter vowel durations, as compared to cases without creaky phonation (cf. Machač and Skarnitzl, 2009a: 131), which was not desirable in this study.

All in all, it is clear that the segmentation of spontaneous speech material is an especially challenging task, and due to a large variability in production patterns, complete consistency of the annotations is difficult to achieve.

In addition to the segmentation, a few other annotations and qualitative evaluations were made in order to describe the tokens as precisely as possible. The immediate context in each token was also registered and marked in the textgrids³⁸ in order to determine the influence of certain context features on the realisation of the function words. In the following three sections, the annotations will be described separately for each function word.

3.2.3.1 Annotations for the word *in*

Figure 3.1 shows an example of segmentation and context-related annotations labelled in the preposition *in*. Tier 1 (context) shows the orthographic transcription of the displayed speech fragment. On tier 2 (word), the observed function word was annotated in the second interval. Tier 3 (segment) contains the annotation of the vowel (v) and nasal (n) segments. In addition, the degree of nasalisation of the vowel was evaluated on a scale from 1 to 3 (1 = not or only slightly nasalised, 2 = moderately nasalised, 3 = heavily nasalised) and marked in the interval corresponding to the vowel segment on tier 4 (sub). The criteria for the evaluation included mainly the auditory impression, comparison with other vowel segments produced by the same speaker and visual cues. As soon as there was a clearly visible and rather abrupt transition between the nasalised vowel and the following nasal segment, the nasalisation degree was not evaluated higher than medium (2). In cases where the vowel nasalisation degree was changing considerably within the vowel segment, two intervals within the vowel were annotated on tier 4, with different nasalisation values assigned to each. Using these values and duration of the respective intervals, one overall value was calculated as an average of

³⁸ Textgrids are Praat-specific annotation files constituted of a desired number of annotation tiers which contain timestamps and labels.

the determined nasalisation values weighted by the duration of the intervals. The resulting value intends to provide a reasonably precise auditory-based evaluation of a quality that is not directly measurable from the acoustic signal. A similar evaluation of the degree of nasalisation was made for the segment immediately following the nasal, and marked to the corresponding interval on tier 4. As opposed to the vowel nasalisation, here we were primarily interested in the amount of nasalisation in the initial portion of the segment, rather than in the precise degree of nasalisation averaged over the whole segment duration. Therefore, only one value on the scale from 1 to 3 was assigned to the segment, even though the nasalisation degree was changing (decreasing) considerably throughout the segment duration. This evaluation was considered irrelevant for nasals and most plosives produced in a usual way. The reason for considering this evaluation irrelevant in plosives is that their mechanism of production (i.e. achieving a complete closure) does not allow for much variation in nasalisation.

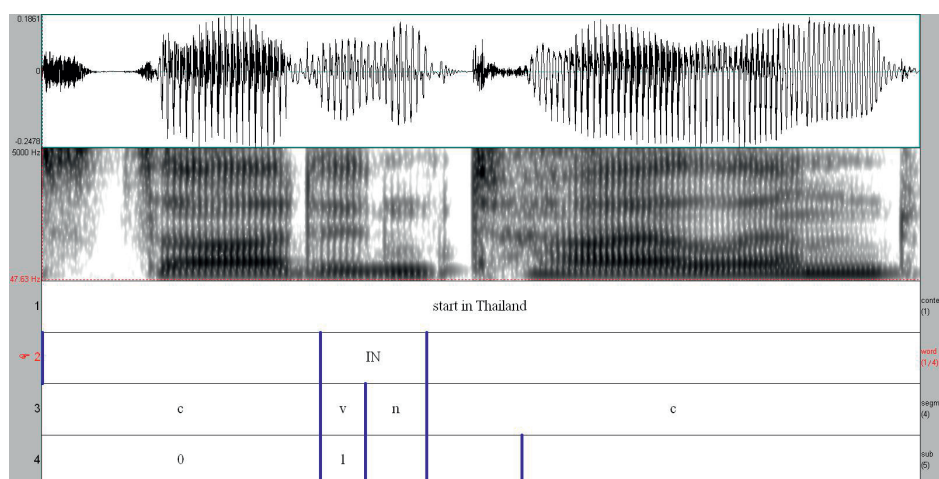


Figure 3.1: An example of segmentation and other annotations labelled in the preposition *in* (for details see text). Tier 1 shows the orthographic transcription of the displayed speech fragment.

As to the neighbouring context, we first classified the segments immediately neighbouring with the observed words on the left and on the right as a consonant (c) or vowel (v), and registered the corresponding code in the first and last interval on tier 3.

Chapter 3

For the preposition *in*, we also registered the phonological identity of the immediately preceding segment. The status of the sound as nasal was marked in the first interval in tier 4 (1 = nasal, 0 = other sound). In addition, we paid attention to the realisation of the segment following the preposition *in*. Here, we frequently encountered a situation, where the segment following the preposition *in* was assimilated to the preceding nasal. For example, this would result in a realisation of the words *in them* as [in:ə̃m]. In such cases, it was not possible to separate the part of the nasal representing the underlying /ð/, and the whole duration of nasal was segmented as part of the preposition *in*. Therefore, we marked such cases as “assimilated” (a) in the interval for nasal segment on tier 4, to be able to treat these cases separately in the subsequent analyses. In some cases, we observed an epenthetic vocalic element following the nasal in the preposition *in*, occurring even in positions where it was obviously not only a result of the movements of the articulators to achieve the position for the next segment. In our segmentation, we separated this segment from the preposition, but we marked these cases as “epenthetic” (e) in the interval for nasal segment on tier 4, assuming that their temporal organisation may deviate from that of items without any inserted vocalic element.

3.2.3.2 Annotations for the word *of*

Figure 3.2 shows an example of segmentation and context-related annotations labelled in the preposition *of*. Tier 1 (context) shows the orthographic transcription of the displayed speech fragment. On tier 2 (word), the observed function word was annotated in the second interval. Tier 3 (segment) contains the annotation of the vowel (v) and fricative (f) segments. In addition, the voiced portion of the vowel (vv) and voiced portion of the fricative (vf) were annotated on tier 4 (sub). The annotations of the context included classification of the sounds immediately neighbouring with the observed word as a consonant (c) or vowel (v). These codes were annotated in first and last interval on tier 3. In the corresponding intervals on tier 4, we annotated the status of the sounds as phonetically voiced (1) or voiceless (0).

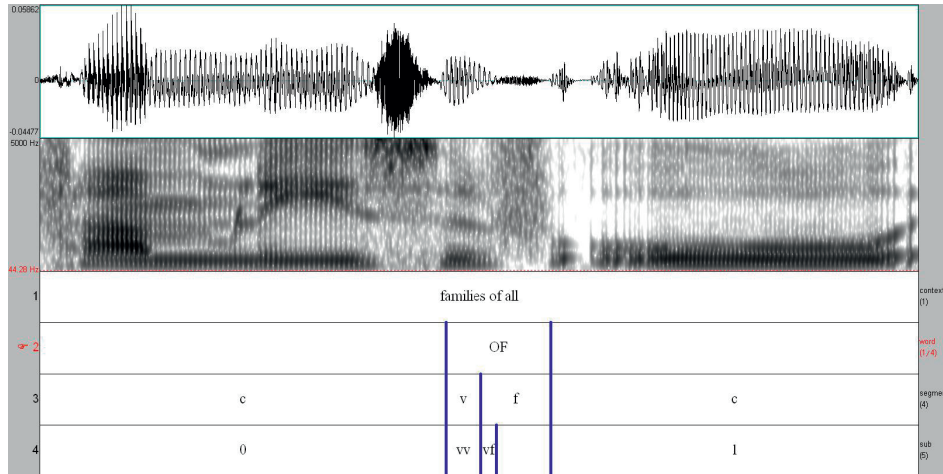


Figure 3.2: An example of segmentation and other annotations labelled in the preposition *of* (for details see text). Tier 1 shows the orthographic transcription of the displayed speech fragment.

3.2.3.3 Annotations for the word *to*

Figure 3.3 illustrates the segmentations and context-related annotations made in the preposition *to*. Tier 1 (context) shows the orthographic transcription of the displayed speech fragment. On tier 2 (word), the observed function word was annotated in the second interval. Tier 3 (segment) contains the annotation of the plosive (p) and vowel (v) segments. In addition, the consonantal closure (cl) and release (r) phases were annotated on tier 4 (sub). The annotations of the context included classification of the sounds immediately neighbouring with the observed word as a consonant (c) or vowel (v). These codes were annotated in the first and last interval on tier 3. In the corresponding intervals on tier 4, we annotated the status of the sounds as phonetically voiced (1) or voiceless (0).

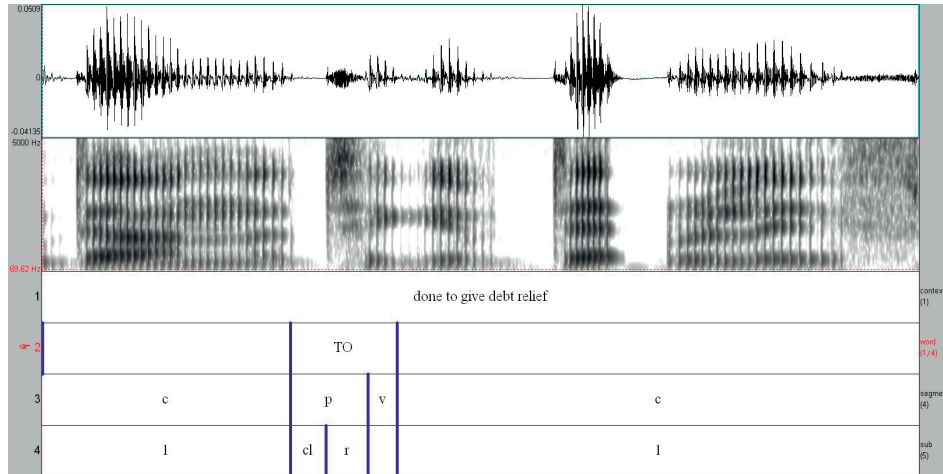


Figure 3.3: An example of segmentation and other annotations labelled in the preposition *to* (for details see text). Tier 1 shows the orthographic transcription of the displayed speech fragment.

3.2.3.4 Articulation rate estimate

Speech rate (or related measures, such as articulation rate) is a factor that has not only an obvious association with word durations, but has also been shown to relate to other aspects of speech reduction (see literature overview, especially Section 1.2.1). To be able to estimate the local articulation rate in the observed tokens, additional annotations were needed. In all speech fragments containing the selected tokens, we annotated portions of clean speech (s), excluding silent pauses and easily separable disfluencies on tier 5 (activity), and marked syllable nuclei within these clean speech intervals as points on tier 6 (syllables). This allowed us to calculate a local articulation rate estimate as the number of syllables divided by clean speech duration in seconds³⁹. Figure 3.4 illustrates the annotation of the number of syllables and speech activity (in this simplified figure, annotations of speech activity and syllable nuclei appear as tiers 3 and 4).

³⁹ Within the the speech fragment containing approximately 1 second of context on each side of the observed function word.

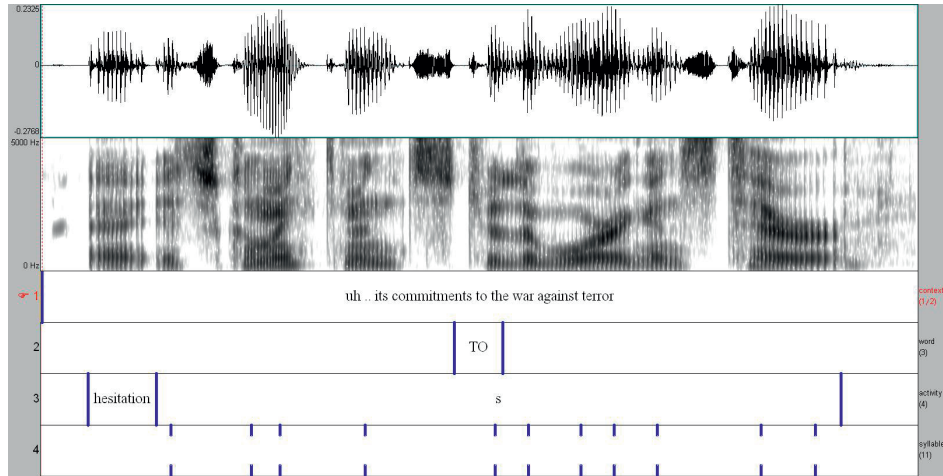


Figure 3.4: An example of annotation of speech activity (on tier 3) and marks for syllable nuclei (on tier 4) (for details see text). Tier 1 shows the orthographic transcription of the displayed speech fragment.

It has to be noted that such an estimate is somewhat less accurate, because it takes into consideration only the few syllables present in the immediate surroundings of the observed word, without regard to their structure. The resulting articulation rate estimate may therefore be biased in some cases, due to the unbalanced distribution of different types of syllables in the actual sample. Clearly, expanding the speech stretch used for the calculation of the articulation rate would increase the number of syllables and their variability in the sample, thus reducing the probability of such bias. With this approach, however, the calculated value would not refer to the close surroundings of the observed word anymore, and it would instead report on a more global articulation rate. We assumed, however, that even though we used the short speech stretches, the large number of tokens used in the study guarantees a low probability of systematic errors. In addition, measuring speech rate in syllables per second also has certain advantages. For instance, Fosler-Lussier and Morgan (1999) point out that this measure of speech rate in spontaneous speech is more stable than a measure of phones per second, because syllables are less likely than phones to be deleted in production. We concluded that for our purposes, the measure of syllables per second used for a large number of small samples is adequate.

3.2.3.5 Acoustic measurements

The annotations of the boundaries of the function words, their segments and subsegmental properties allowed us to perform automatic measurements of durations. We also calculated the relevant segment ratios (e.g. proportions of the duration of the vowel within the whole item duration, or proportions of the release phase duration within the plosive duration in the word *to*) in the observed words. All these proportions were expressed as percentages. Apart from that, we obtained a number of other measurements. In order to inspect the vowel quality, the formant values (F1, F2 and F3) in Bark were measured in the words *of* and *to*, as the means of values obtained from the whole duration of the vowel in the observed item. No similar formant measurements were made in the word *in*, because the degree of nasalisation of the vowel and its segmental context were expected to have a major influence on the formant values, making such measurements of little value. To be able to eliminate the relatively frequent errors in automatic formant tracking in Praat, an additional semi-automatic method was used to detect any abrupt jumps between nearby measurements of the same formant, as determined by the Praat formant-tracking algorithm. The items where a jump of over 3 Bark within 12.5 ms was present in the formant measurements were then manually checked, and the automatically obtained values were corrected to match the formants as determined by a careful inspection of the spectrogram. In a few cases, where the vowel energy was too low, and the formant measurements were therefore unreliable, the formant values were discarded. The resulting formant values were used to calculate the F1 - F0 and F3 - F2 values in Bark (*Bark distances*), as measures of vowel openness and backness, respectively (Syrdal and Gopal, 1986). According to Syrdal (1985) and Adank et al. (2004), this normalisation method successfully reduces gender-related variation in the data. The value of F0 necessary for this transformation was measured in the centre of the vowel interval, avoiding the portions with a creaky voice quality where possible, and converted to Bark using Traunmüller's formula (Traunmüller, 1990).

In addition, we used an acoustic measure of *friction intensity* that refers to the phonological voicing (fortis/lenis) distinction in obstruents. It is known that as a result

of the production mechanism of obstruent sounds, for a given air pressure, the acoustic intensity of the friction noise component of voiced obstruents is lower than that of corresponding voiceless sounds (e.g. Halle et al., 1957; Strevens, 1960; Smith, 1997). The balance between the spectral energy in low and high frequencies may therefore be used as another measure of the fortis/lenis character of obstruents. For that purpose, band energies (low-frequency band: 0 to 5000 Hz, high-frequency band: over 5000 Hz) were measured in the fricative, and in the release portion of the plosive in words *of* and *to*, respectively. The band energies were then used to calculate the *high-frequency band - low-frequency band* differences in dB, corresponding to relative friction intensity. In the case of the preposition *of*, this measure complements the duration-based measure of the phonetic voicing proportion within the fricative, although a rather high correlation between the two measures may be expected.

3.2.4 Praat scripts

After all the annotations were prepared in the textgrids, the durations as well as other data were extracted automatically using Praat scripts. This approach, apart from saving a considerable amount of time, allowed for various checks of the consistency of the annotations. The main output consisted of tables (text files) containing all the data ready for use in the statistic software.

3.2.5 Statistical analyses

Most of the described statistical analyses were carried out using the statistical software SPSS for Windows, version 15.0 (SPSS Inc., 2006). The statistical tests using an analysis of variance were performed with the Type III Sums of Squares, which is recommended in most research settings, including unbalanced designs (e.g. Iacobucci, 1995). When Levene's test showed heterogeneity of variances at a level of significance $p < 0.001$, a series of Kruskal-Wallis tests were carried out instead of an analysis of variance. The non-parametric statistical tests were carried out using the statistical software R (R Development Core Team, 2010).

3.3 Results

In this section, we will describe the results, presenting first several global analyses relating mainly to the influence of the context on word realisations (Section 3.3.1), followed by detailed analyses for each of the three function words (Sections 3.3.2 – 3.3.4). Section 3.4 will then summarise and discuss the main context effects on the function word realisations and the results found for the three words.

3.3.1 Context influence

3.3.1.1 Articulation rate

We measured the articulation rate (in syllables per second) in speech fragments surrounding the observed words (see Section 3.2.3.4 for details) in order to establish its contribution to the variability in function word realisations. Figure 3.5 shows the boxplots of the articulation rates for groups of items produced by speaker groups with different L1 backgrounds in the two speaking styles (measured in the speech fragments surrounding the three different function words). The boxes in the plots contain the inter-quartile range, and the whiskers extend to the furthest datapoint that lies within 1.5 times the inter-quartile range from the box on each side.

In total, we observed higher articulation rates in the native speaker group (5.78 syllables per second) and slower articulation rates for the non-native speakers (5.04 syll/s for the Czech speakers, and 5.42 syll/s for the Norwegian speakers). We also observed consistently higher articulation rates in read speech than in spontaneous speech. Pooled over the different speaker groups, the mean articulation rate was 5.52 syll/s in read speech, and 5.30 syll/s in spontaneous speech. The means and other statistical indicators of the groups of items based on the speakers' L1 background and speaking style are listed in Table 3.1 below.

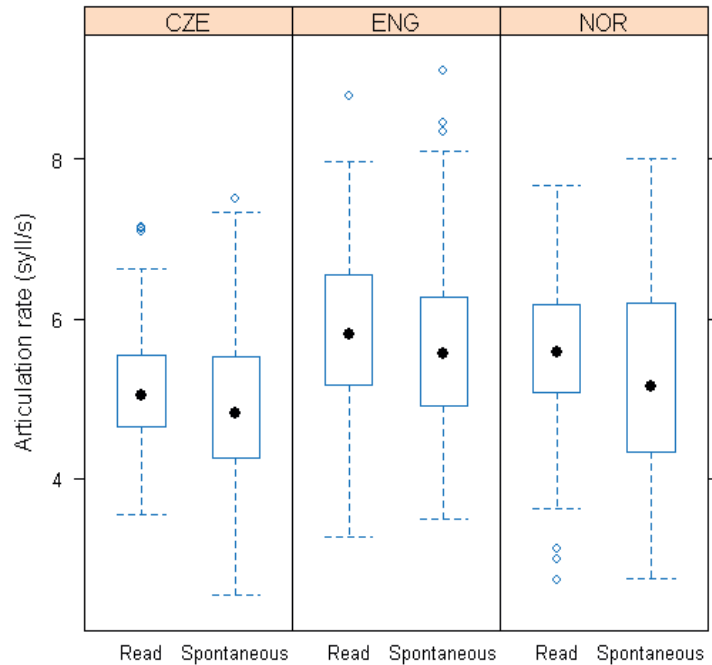


Figure 3.5: Articulation rates (in syllables per second), for speaker groups with different L1 backgrounds (CZE=Czech, ENG=English, NOR=Norwegian) in the two speaking styles.

Since Levene’s test indicated unequal variances with high significance ($p < 0.001$), we performed the statistical tests using non-parametric tests. Kruskal-Wallis tests showed a significant effect of both L1 ($\chi^2(2) = 80.0$; $p < 0.001$) and speaking style ($\chi^2(1) = 12.1$; $p < 0.001$). A post-hoc test using Mann-Whitney tests with Bonferroni correction showed that the Czech speaker group differed significantly both from the native English speakers ($p < 0.001$) and from the Norwegian speakers ($p < 0.001$), and the group of Norwegians still differed significantly from the natives ($p = 0.002$). In addition, we inspected the articulation rate variability as described by standard deviations. Table 3.1 lists the means, standard deviations and corresponding variation coefficients of the articulation rates measured in the speech fragments surrounding the observed words, for the groups of items based on speakers’ L1 and speaking style. A variation coefficient is

Chapter 3

a standard deviation normalised against the group mean, expressed as a percentage. It can be observed that for all three speaker groups (based on the speakers' L1 background) the variability of the observed articulation rates is higher in spontaneous speech. This is especially noticeable in Norwegian items.

Table 3.1: Means (in syll/s), standard deviations (in syll/s) and variation coefficients (in %) of articulation rates measured in speech fragments surrounding the observed words, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

	mean (syll/s)	std. dev. (syll/s)	var. coeff. (%)
CZ-R	5.11	0.72	14
CZ-S	4.96	0.92	19
EN-R	5.86	0.94	16
EN-S	5.68	1.02	18
NO-R	5.59	0.89	16
NO-S	5.24	1.24	24

Table 3.2 lists the overall values for the three observed speaker groups and for the two speaking styles. The variation coefficient of the Norwegian items is somewhat higher than in the Czech and native English group. This is probably due to their large variability in spontaneous speech. As for the speaking styles, the overall variation coefficients were 16% for the read items, and 21% for the spontaneous items. This indicates noticeably greater articulation rate variability in spontaneous speech.

Table 3.2: Overall means (in syll/s), standard deviations (in syll/s) and variation coefficients (in %) of articulation rates measured in speech fragments surrounding the observed words, for the three speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and the two speaking styles.

	mean (syll/s)	std. dev. (syll/s)	var. coeff. (%)
CZ	5.04	0.82	16
EN	5.78	0.98	17
NO	5.42	1.09	20
Read	5.52	0.91	16
Spontaneous	5.30	1.11	21

To determine how much of this variability is due to the individual speakers' increased articulation rate variability in spontaneous speech, as opposed to the mere increase of inter-speaker differences in spontaneous speech, we calculated the variation coefficients for each speaker's read and spontaneous items and compared these for each speaker. Twenty four out of the 30 speakers had higher articulation rate variation coefficients in spontaneous speech than in read speech (see Figure 3.6).

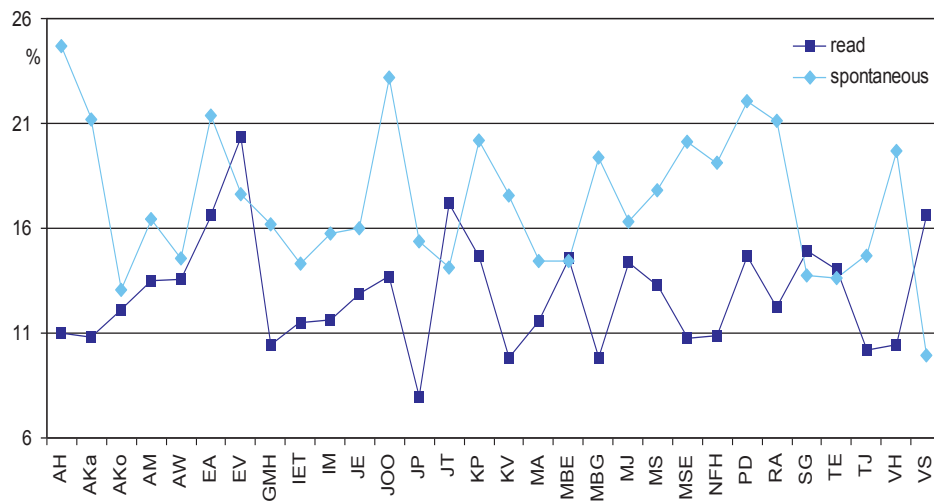


Figure 3.6: Variation coefficients (in %) of articulation rates for each speaker's read and spontaneous items. Speakers' initials are sorted alphabetically.

As mentioned previously, speech rate is a factor that has a clear association with word durations. Previous research has, however, also shown its relation to other aspects of speech reduction. We inspected the correlation of articulation rate with word duration for each of the three function words, and revealed moderate negative correlations for the words *in* and *to* ($r = -0.479$; $N = 289$; $p < 0.001$ and $r = -0.495$; $N = 285$; $p < 0.001$, for the words *in* and *to*, respectively) and a slightly weaker negative correlation for the word *of* ($r = -0.367$; $N = 294$; $p < 0.001$). Scatterplots in Figures 3.7, 3.8 and 3.9 illustrate the relationship of articulation rate and item duration in the realisations of the words *in*, *of* and *to*, respectively.

Chapter 3

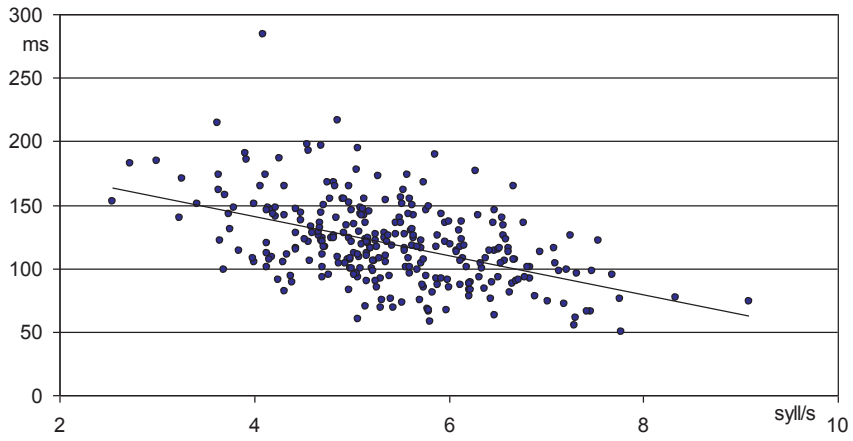


Figure 3.7: Relationship of articulation rate (in syll/s) and item duration (in ms) in the realisations of the function word *in* ($r = -0.479$; $p < 0.001$; $b = -15.33$).

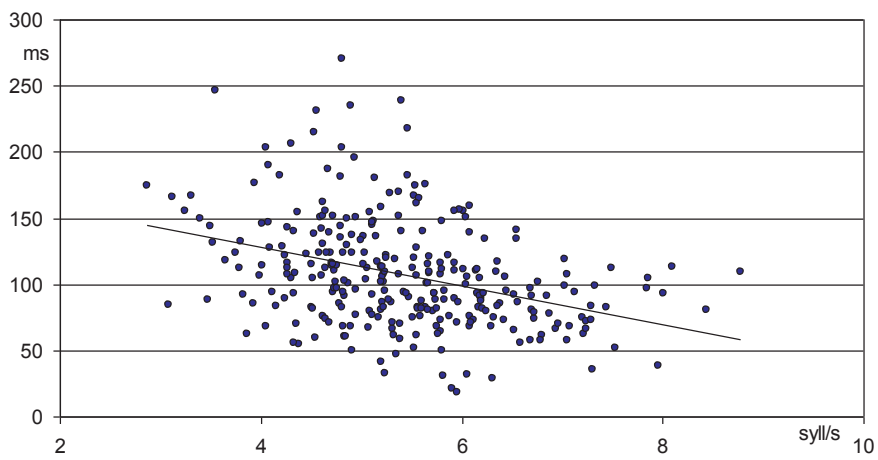


Figure 3.8: Relationship of articulation rate (in syll/s) and item duration (in ms) in the realisations of the function word *of* ($r = -0.367$; $p < 0.001$; $b = -14.66$).

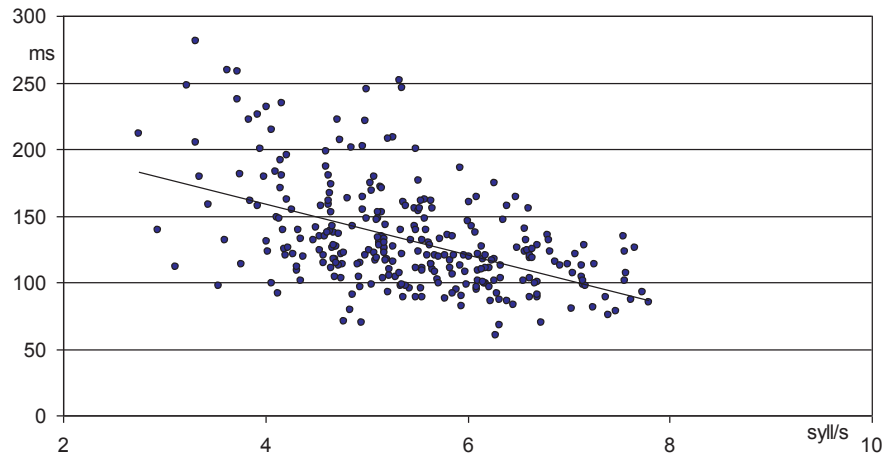


Figure 3.9: Relationship of articulation rate (in syll/s) and item duration (in ms) in the realisations of the function word *to* ($r = -0.495$; $p < 0.001$; $b = -18.95$).

Because of these correlations and the previously determined significant differences in the mean articulation rate depending on the L1 background and speaking style, it was not considered meaningful to carry out further analyses of the factors affecting item durations without taking the articulation rate into consideration. To avoid using the strongly articulation rate-dependent raw durations in further analyses, we calculated normalised duration values “freed” from articulation rate influence. The normalised values are intended to express the relative durations of function words, with respect to the overall speech production, rather than the actual durations, enabling us to show differences that are not due to the articulation rate only.

In order to achieve an appropriate normalisation, it was necessary to describe the relationship of articulation rate and item durations precisely. It was obvious that the relationship of articulation rate and item durations may not be identical for the different words. Therefore, we determined the regression coefficients characterising the relationship of articulation rate and item duration separately for the three words. The different relationships of articulation rate and item duration for the three function words can be observed from the scatterplots in Figures 3.7, 3.8 and 3.9. Table 3.3 lists the

Chapter 3

correlation coefficients and correlation significance levels, regression slopes and mean articulation rates for each of the three words. Subsequently we used the articulation rate means and regression slopes to transform the raw durations into normalised durations, using a formula:

$$\text{Dur}_{\text{norm}} = \text{Dur}_{\text{raw}} - (\text{Slope} * (\text{Rate}_{\text{act}} - \text{Rate}_{\text{mean}})) ,$$

where Dur_{norm} is the resulting normalised duration, Dur_{raw} is the original (raw) item duration, Slope is the regression slope for the given function word, Rate_{act} is the actual articulation rate at which the given item was produced, and $\text{Rate}_{\text{mean}}$ is the mean articulation rate for the given function word. The resulting normalised duration values were larger than the raw duration values in tokens with higher (actual) articulation rates, or smaller than the raw durations, in tokens produced at low articulation rates. These normalised word duration values were used in all further analyses instead of the raw measured durations, to ensure that the variance due to articulation rate is minimised.

Table 3.3: Correlation coefficients and significance, regression slopes and articulation rate means (in syll/s) describing the relationship of articulation rate and item duration, for the three observed function words.

	r	p	slope	a. r. mean
IN	-0.479	< 0.001	- 15.33	5.41
OF	-0.367	< 0.001	- 14.66	5.41
TO	-0.495	< 0.001	- 18.95	5.41

As for the influence of articulation rate on other observed variables (e.g. segment proportions, formant values), all the correlations tested separately on data from each function word were shown to be very low (mostly around $|r| = 0.1$). Even though some of these correlations reached significance, the low correlation coefficients imply a low inter-variable determination. Therefore, the influence of articulation rate on all the variables except item duration was considered irrelevant.

3.3.1.2 Segmental context

In order to control for the possible effects of neighbouring segments (cf. Section 1.2.1.1), we classified the segments immediately neighbouring with the observed words as a consonant or vowel. The segment immediately preceding the observed function word will also be called *left segment* and the segment immediately following the function word will be called *right segment*. Firstly, we inspected the distributions of the neighbouring segment types in the data. Table 3.4 lists the numbers of tokens in the subgroups split up based on word, speakers' L1 background and speaking style that had different types of segment (consonant or vowel) as their left segment. Similarly, Table 3.5 presents the numbers of tokens in the defined subgroups having different types of segment (consonant or vowel) as their right segment.

Table 3.4: Numbers of tokens where the observed word is preceded by different types of segment (consonant or vowel), for the subgroups of tokens split up based on word, speaker L1 and speaking style.

	IN		OF		TO	
	consonant	vowel	consonant	vowel	consonant	vowel
CZ-R	47	3	47	3	39	11
CZ-S	38	6	42	4	37	6
EN-R	44	6	46	4	41	8
EN-S	40	8	47	2	28	17
NO-R	44	6	42	8	42	8
NO-S	44	3	48	1	39	9

Table 3.5: Numbers of tokens where the observed word is followed by different types of segment (consonant or vowel), for the subgroups of tokens split up based on word, speaker L1 and speaking style.

	IN		OF		TO	
	consonant	vowel	consonant	vowel	consonant	vowel
CZ-R	40	10	38	12	45	5
CZ-S	36	8	41	5	40	3
EN-R	34	16	34	16	44	5
EN-S	39	9	39	10	44	1
NO-R	39	11	39	11	40	10
NO-S	34	13	41	8	45	3

Chapter 3

Pooled across the word identity, the distribution of consonants and vowels in the neighbouring segments does not seem to differ much between the groups based on speaker L1 and speaking style. The following two figures show the distributions of segment types preceding (Figure 3.10) and following (Figure 3.11) the observed words. Pearson's Chi-squared tests showed that neither for the left segment type nor the right segment type distributions, was there a significant difference between the groups of items based on the speakers' L1 background and speaking style ($\chi^2(5) = 7.50$; $p = 0.186$ and $\chi^2(5) = 10.7$; $p = 0.057$, for left and right segment type, respectively).

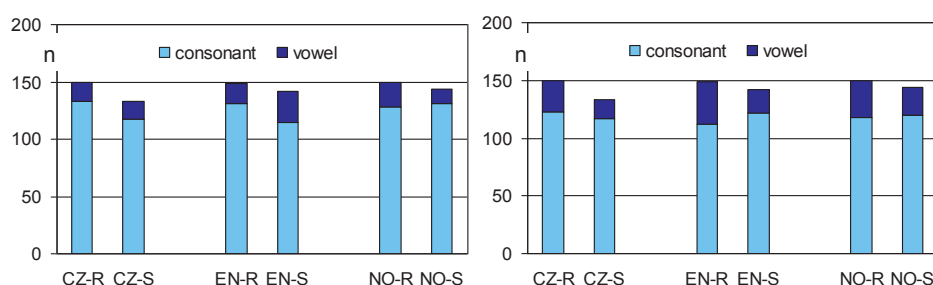


Figure 3.10 (on the left) and Figure 3.11 (on the right): Distributions of left (fig. 3.10) and right (fig. 3.11) segment types (in numbers of tokens) in items pooled across the three function words, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

However, when we inspect the distributions of consonants and vowels in the neighbouring segments for the three function words, pooling across the speaker L1 background and speaking style (see Figures 3.12 and 3.13), it is immediately obvious that the differences are considerable. Pearson's Chi-squared tests showed that the three words had significantly different distributions of both left segment type ($\chi^2(2) = 23.8$; $p < 0.001$) and right segment type ($\chi^2(2) = 21.2$; $p < 0.001$). The different distributions of segment types in the neighbourhood of the three different function words may be partly explained by the frequency of certain word combinations in English.

Although it would be interesting to analyse any global influence of context type on temporal variables, such as normalised word duration or vowel proportion, it turned out

to be technically complicated. Since Levene’s test showed a highly significant inequality of variances, the analyses would have to be carried out using Kruskal-Wallis tests, without the possibility to use multifactorial design and investigate the interactions of various factors. At the same time, interactions of context type with the factor word can be expected, since the three words have considerably different structure. Therefore, we decided to perform the analyses of context type influence separately for each word, and they will be described together with other context-related analyses (e.g. influence of phonetic voicing in a neighbouring segment) in the following sections.

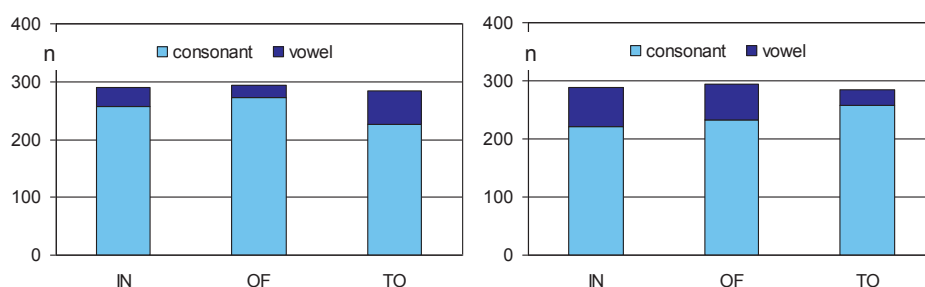


Figure 3.12 (on the left) and Figure 3.13 (on the right): Distributions of left (fig. 3.12) and right (fig. 3.13) segment types (in numbers of tokens) in items pooled across speaker L1 background and speaking style, for the three function words.

Moreover, it has to be kept in mind that although we attempt to describe the influence of context on the investigated variables, the design of this study does not allow us to make definitive conclusions. Unlike studies using laboratory materials with full control over all the context parameters, our sample collected from natural productions in two speaking styles constitutes an unbalanced design. Although we may reveal tendencies resulting from the strong effects of some of the observed context factors, the present design does not guarantee that the factors are completely independent, and the effects of individual factors may not be reliably isolated. To be able to reliably determine the effects of individual context factors, a balanced study with systematically varied context parameters would be necessary. In addition, this study inspected only a few of the possibly influential context parameters, while many others, such as prosodic

prominence or place of articulation of neighbouring segments, were not considered at all. The scope of this analysis of context effects is, therefore, inherently limited.

3.3.2 IN

3.3.2.1 *Temporal organisation in the word in*

As for the temporal organisation in the preposition *in*, we first inspected the normalised whole word durations of the items. Figure 3.14 shows boxplots of the normalised durations of the word *in* produced in read and spontaneous speech for the three speaker groups based on their L1 background. The boxes in the plots contain the inter-quartile range, and the whiskers extend to the furthest datapoint that lies within 1.5 times the inter-quartile range from the box on each side.

The normalised durations were on average 120 ms for the Czech speakers, 119 ms for the native English speakers and 119 ms for the Norwegian speakers, indicating that after normalising for the effect of articulation rate, the differences in duration between the different speaker groups with different L1 backgrounds are rather slight. In addition, we found out that the values were slightly shorter in read speech (116 ms) than in spontaneous speech (122 ms). More detailed statistical information relating to normalised word durations, as well as to other temporal and spectral variables inspected in this chapter, are listed in Appendix B.

An analysis of variance with the factors L1 background and speaking style showed that neither of the two factors was significant ($F(2, 283) < 1$ and $F(1, 283) = 3.22$; $p = 0.074$). The interaction between the two factors did not reach significance either ($F(2, 283) < 1$).

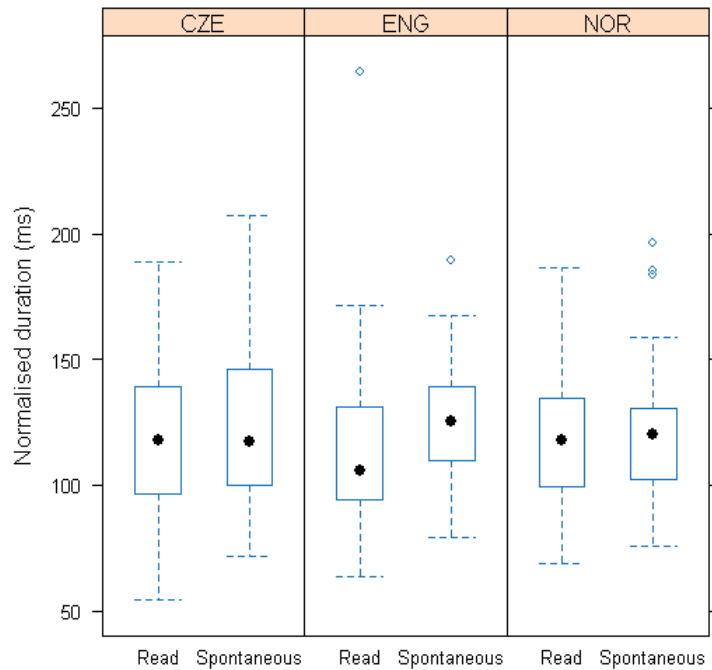


Figure 3.14: Normalised word durations (in ms) in the word *in*, for speaker groups with different L1 backgrounds (CZE=Czech, ENG=English, NOR=Norwegian) in the two speaking styles.

To study the internal syllable structure we calculated the segment ratios, i.e. the proportions of the duration of the vowel within the whole item duration, and used these percentage values to calculate the means for individual groups of items. It is important to mention that this percentual ratio is not precisely equal to the overall ratio that we could obtain by calculating the ratio of the mean vowel duration within the mean word duration. Such a value would then express a weighted average of segment ratios (weighted by item duration) rather than an average where each item counts the same, regardless of its duration. The mean vowel proportion across the speaker groups and speaking styles was 45%, and it did not vary much between the speaker groups (46% for Czech speakers, 44% for English speakers and 47% for Norwegians) and between the speaking styles (45% in read speech, and 46% in spontaneous speech). An analysis

of variance confirmed these findings, showing no significant effect of either speaker group or speaking style ($F(2, 283) = 1.28$; $p = 0.280$ and $F(1, 283) < 1$). There was no significant interaction between the two factors, either ($F(2, 283) = 1.29$; $p = 0.277$).

3.3.2.2 Vowel nasalisation in the word in

The only variable relating to the quality of the vowel in the preposition *in* was the auditory impression of the degree of nasalisation. The nasalisation degree was evaluated using a scale: 1 (not or only slightly nasalised), 2 (moderately nasalised) or 3 (heavily nasalised), but the overall evaluation of a vowel could assume any value between 1 and 3, since vowels containing portions with varying quality were assigned separate values for the different portions, and a weighted average of these values was calculated. The mean values varied slightly between the three speaker groups, with the mean nasalisation degree of 1.85 for the Czech speakers, 1.91 for the native English speakers and 1.97 for the Norwegian speakers. The means for different speaking styles were almost equal (1.92 and 1.91 for read and spontaneous speech, respectively). An analysis of variance confirmed this observation, showing no significant effect of either speaker group or speaking style ($F(2, 281) = 1.07$; $p = 0.345$ and $F(1, 281) < 1$). There was also no significant interaction between the two factors ($F(2, 281) < 1$).

3.3.2.3 Interaction of the word in with context

We will first describe the distribution of special cases, where the preposition *in* was produced either with the following segment assimilated to the nasal, or with an epenthetic vocalic segment inserted between the preposition and the following segment (see Section 3.2.3.1 for details). The numbers of both types of these cases were rather low, amounting to 8 cases of assimilated segments and 3 cases of epenthetic segment insertions in a total of 289 tokens. The segment assimilations were produced mainly by the Norwegian speakers (6 cases), and English speakers (2 cases), and all of them occurred in spontaneous speech. It has to be noted that this can be partly ascribed to the very frequent occurrence of the definite article (*the*) directly following the preposition.

In spontaneous speech, the definite article followed the preposition in 81 cases out of 139 tokens, while in read speech it occurred in this position only 38 times out of 150 tokens. The assimilation of the initial fricative in *the* is facilitated due to the similar place of articulation, and the status of the article as one of the most frequent words certainly does not call for increased articulation precision. In read texts, however, the proportion of other words starting with vowels, plosives or other sounds that were not so strongly predisposed to the assimilation, was much higher. Moreover, this phenomenon has been observed to occur often in casual speech (cf. Manuel, 1995; Shockey, 2003: 43-44). This would be consistent with a higher incidence in spontaneous speech compared to the presumably more careful read speech in our data. Unlike the occurrences of assimilated segment, the cases containing an epenthetic vocalic sound were produced exclusively by the Czech speakers in read speech. The tendency to insert epenthetic schwa was observed in Czech in sequences of consonants with the same manner of articulation (Machač and Skarnitzl, 2009a: 116-119, 122-123), and following pre-pausal speech sounds or on word boundaries, where the schwa reinforces the preceding speech sound (Matoušek et al., 2009). Other studies (Machač and Skarnitzl, 2009b; Skarnitzl and Machač, 2012) reported a frequent occurrence of this phenomenon in the speech of Czech television and radio broadcasters, and indicated its possible association with affectedness or exaggerated effort to speak in a correct way. Since such stylisations are characteristic of media professionals' performance, some Czech speakers may tend to unintentionally imitate the particular phenomena when focussing on a task such as news reading. Due to the small number of these special cases in our material, it was not possible to perform any statistical analyses on these subsets of data. To reveal whether including these special cases caused any distortion of the results, we repeated the analyses of variance for the main effects, excluding these items from the data. As expected, the results did not differ from the results described above.

Another aspect of the interplay between the realisation of the preposition *in* and its context is the possibility of the nasalisation of segments following the nasal in the word *in*. To examine this coarticulatory phenomenon, we evaluated the degree of nasalisation

Chapter 3

in the segment immediately following the word, whenever this was relevant (see Section 3.2.3.1 for details). Figure 3.15 shows the mean values of nasalisation degree in the following segment, for speaker groups with different L1 backgrounds and different speaking styles. The mean values varied between the three speaker groups with the mean nasalisation degree of 1.72 for the Czech speakers, 2.20 for the English speakers and 2.07 for the Norwegian speakers. The mean degree of nasalisation of the following segment was 1.94 in read speech and 2.07 in spontaneous speech. Since the measure of nasalisation degree in the segment following the preposition *in* assumed only the values 1, 2 or 3, it needs to be treated as an ordinal variable, and non-parametric tests have to be used. A Kruskal-Wallis test revealed a significant effect of L1 on the nasalisation of the following segment ($\chi^2(2) = 18.7$; $p < 0.001$). A post-hoc test using Mann-Whitney tests with Bonferroni correction showed that the Czech speaker group differed significantly both from native English speakers ($p < 0.001$) and from Norwegian speakers ($p = 0.009$). The effect of speaking style did not reach significance ($\chi^2(1) = 2.03$; $p = 0.154$). In addition to the main effect of the native language background, we could observe an interaction between the speaker group and speaking style: while both the Czech and native English speakers had almost the same mean values of the following segment's nasalisation degree in read and spontaneous speech (1.74 vs. 1.71 for Czech speakers and 2.22 vs. 2.19 for native English speakers), for Norwegians the mean values for the two speaking styles differed considerably (1.83 vs. 2.31 for read and spontaneous speech, respectively). To test this interaction, we tested the L1 effect for each speaking style separately. The results showed that the L1 effect was significant in both speaking styles ($\chi^2(2) = 9.13$; $p = 0.010$ for read speech and $\chi^2(2) = 16.1$; $p < 0.001$ for spontaneous speech), but there was a difference in paired comparison results. According to Mann-Whitney tests with Bonferroni correction, in read speech, the Czech speaker group differed significantly only from the natives ($p = 0.019$), but also the Norwegian speakers were almost significantly different from the natives ($p = 0.052$). In spontaneous speech, on the other hand, the Czech speakers' values differed significantly from both native speakers' values ($p = 0.004$) and those of Norwegian speakers ($p < 0.001$), and there was no significant difference between the Norwegians and natives.

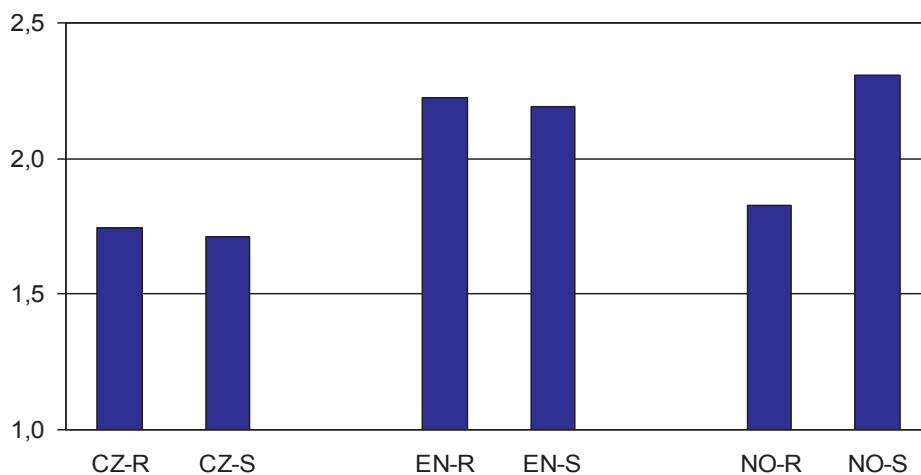


Figure 3.15: Mean values of nasalisation degree (based on auditory evaluation on a scale from 1 to 3) of the segment following the preposition *in*, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

In addition to the above presented analyses of some aspects of context realisation, we intended to analyse in detail the influence of certain aspects of context on the realisations of the preposition *in*. Firstly, we inspected the influence of sound type (consonant or vowel) of the segment immediately preceding or following the observed word on normalised word durations. The mean normalised durations were slightly shorter for words preceded by consonants (118 ms) than for those preceded by vowels (128 ms). There was also a difference between the groups of items followed by consonants and vowels (122 ms and 110 ms, respectively). An analysis of variance with the factors L1, speaking style, left segment type and right segment type revealed only a significant effect of the left (preceding) segment type ($F(1, 267) = 4.86$; $p = 0.028$). Similarly we examined the effect of immediately preceding or following segment type on vowel proportion in the word realisations. The vowel proportion values did not seem to vary much depending on the preceding segment type (mean ratio being 45% in both contexts). The following segment type, however, seemed to influence the vowel ratio considerably, the mean values being 44% and 50% for items followed by consonants and vowels, respectively. An analysis of variance with the factors L1 background,

Chapter 3

speaking style, left segment type and right segment type revealed a significant effect of the right segment type ($F(1, 267) = 4.13$; $p = 0.043$). As for the degree of nasalisation in the vowel or the following segments, we did not consider it relevant to inspect the influence of neighbouring segment types on these measures, and therefore, no results will be reported.

Lastly, we examined the effects of the preceding segments' phonological identity as nasal. In our data, this context variable did not prove to have a significant effect on normalised duration and nasalisation degree in the segment following the observed preposition, but we found significant effects on the other observed variables: proportion of vowel duration within the word and nasalisation degree of the vowel. Figure 3.16 displays the normalised durations of items preceded by nasals compared to cases preceded by non-nasals for groups based on speaker L1 and speaking style.

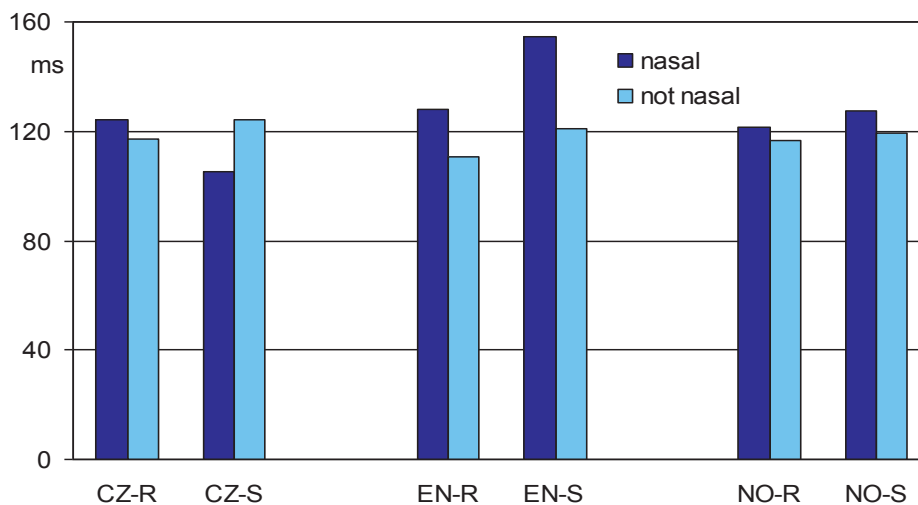


Figure 3.16: Normalised duration (in ms) of the preposition *in*, depending on the phonological identity of the segment preceding the preposition as nasal or non-nasal, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

In sum, the normalised durations of words preceded by a non-nasal were slightly shorter (118 ms) than those of items preceded by a nasal segment (126 ms). An analysis of variance with the factors L1 background, speaking style and the preceding segment's phonological identity as nasal, however, only found a significant interaction of L1 background with the preceding segment's phonological identity ($F(2, 277) = 3.37$; $p = 0.036$). The effect of the preceding segment's identity did not reach significance ($F(1, 277) = 3.39$; $p = 0.067$). As can be seen from Figure 3.16, the interaction is due to a native speakers' more noticeable difference in durations depending on the context, compared to negligible differences in both non-native groups.

Figure 3.17 illustrates the higher proportions of the duration of vowel within word duration for items preceded by nasals, compared to cases preceded by non-nasals (mean vowel proportion pooled across the L1 and speaking styles being 51% vs. 44%, respectively), consistently in all groups based on the speakers' L1 backgrounds and speaking style. An analysis of variance with the factors L1 background, speaking style and preceding segment's phonological identity as nasal showed only the effect of the preceding segment's identity ($F(1, 277) = 8.15$; $p = 0.005$) with no significant interactions.

It was not surprising that the mean degree of nasalisation in the vowel of the preposition *in* also differed depending on the preceding segment's identity as nasal (with overall mean values 2.31 vs. 1.84 for items following nasals and non-nasals, respectively). Furthermore, it seemed that the difference between the nasalisation degree following a nasal and non-nasal segment is more noticeable in items spoken by native speakers. Mean values of nasalisation degree in the vowel in the word *in* depending on the phonological identity of the segment preceding the preposition as nasal or non-nasal for speaker groups with different L1 backgrounds and different speaking styles are displayed in Figure 3.18. An analysis of variance with vowel nasalisation degree as a dependent variable and the factors L1 background, speaking style and preceding segment's phonological identity as nasal, showed only the significant effect of the preceding segment's identity as nasal ($F(1, 275) = 19.2$; $p < 0.001$).

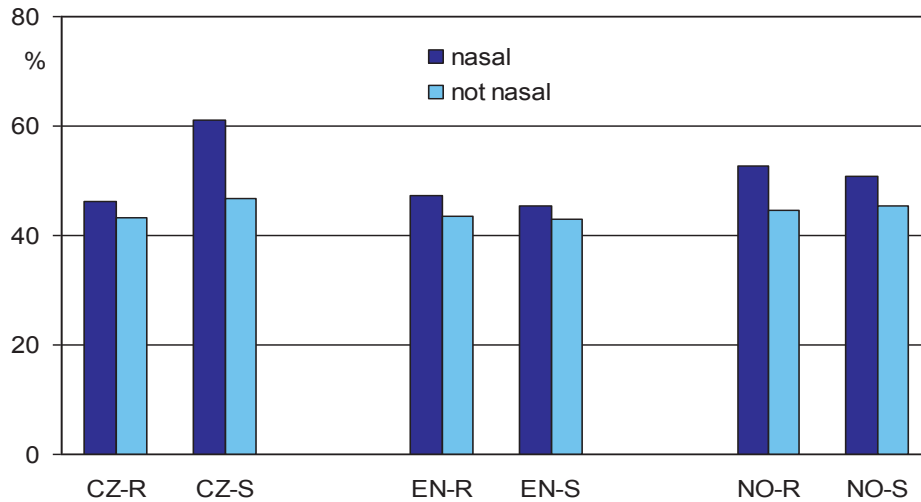


Figure 3.17: Proportion of vowel duration (in %) within the preposition *in*, depending on the phonological identity of the segment preceding the preposition as nasal or non-nasal, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

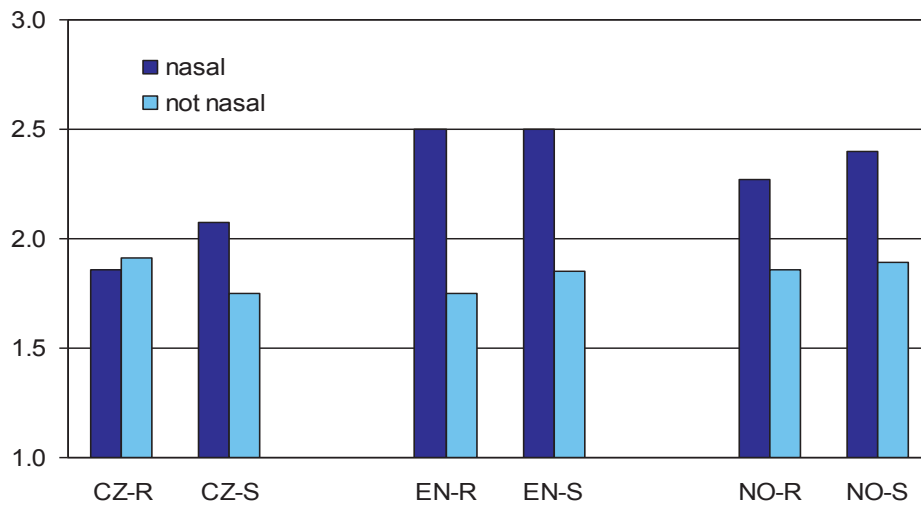


Figure 3.18: Mean values of nasalisation degree (based on auditory evaluation on a scale from 1 to 3) of the vowel in the preposition *in*, depending on the phonological identity of the segment preceding the preposition as nasal or non-nasal, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

In addition to describing the effects of context factors on the realisations of the preposition *in*, we considered it appropriate to inspect the distribution of the various context types in the present dataset. First we inspected the distributions of segment types of the immediately neighbouring left and right segment, comparing the distributions between the groups based on the speakers' L1 backgrounds and speaking style (see Tables 3.4 and 3.5 on page 98). Pearson's Chi-squared tests showed that neither the left segment type distributions nor the right segment type distributions varied significantly between the groups ($\chi^2(5)= 4.26$; $p= 0.512$ and $\chi^2(5)= 4.18$; $p= 0.523$, for left and right segment type distributions, respectively). Likewise, the distributions of phonological nasals among the segments preceding the word *in* did not differ between the subgroups based on L1 and speaking style, as confirmed by a Pearson's Chi-squared test ($\chi^2(5)= 6.09$; $p= 0.298$).

3.3.3 OF

3.3.3.1 Temporal organisation in the word of

As a first measure of the temporal organisation of the function word *of*, we inspected the normalised word durations. Figure 3.19 shows boxplots of the durations of the word *of* in read and spontaneous speech for the three speaker groups based on their L1 background. The boxes in the plots contain the inter-quartile range, and the whiskers extend to the furthest datapoint that lies within 1.5 times the inter-quartile range from the box on each side.

The mean normalised durations were 124 ms for the Czech speakers, 97 ms for the native English speakers and 103 ms for the Norwegian speakers. The values were also longer in read speech (115 ms) than in spontaneous speech (100 ms). Since Levene's test indicated unequal variances with high significance ($p < 0.001$), we performed the statistical tests using non-parametric tests. Kruskal-Wallis tests showed a significant effect of both L1 ($\chi^2(2)= 24.8$; $p < 0.001$) and speaking style ($\chi^2(1)= 6.12$; $p= 0.013$). A post-hoc test using Mann-Whitney tests with Bonferroni correction showed that the Czech speaker group differed significantly both from the native English speakers ($p <$

Chapter 3

0.001) and from the Norwegian speakers ($p= 0.001$). In order to test the interaction of the two factors, we performed Kruskal-Wallis tests separately for read and spontaneous items. The results showed that the L1 effect was significant only in read speech ($\chi^2(2)= 24.0$; $p< 0.001$) but not in spontaneous speech ($\chi^2(2)= 5.53$; $p= 0.063$).

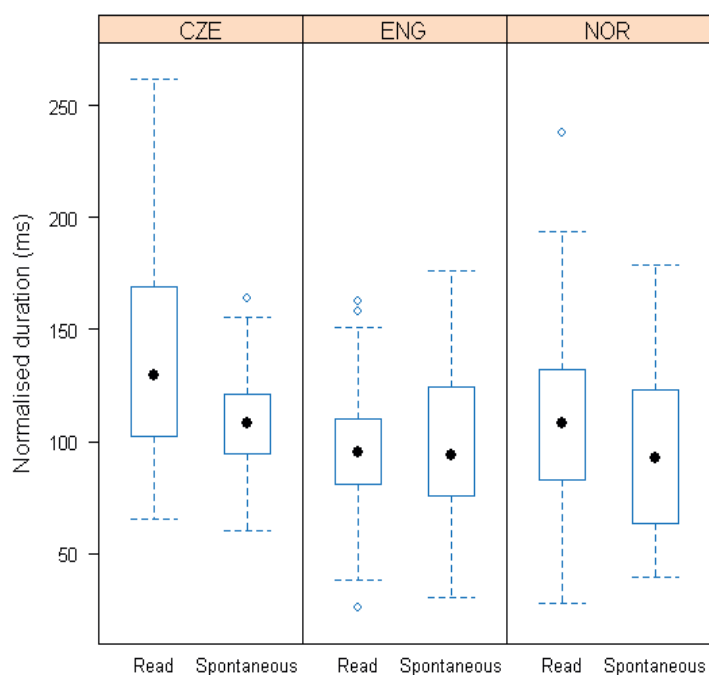


Figure 3.19: Normalised word durations (in ms) in the word *of*, for speaker groups with different L1 backgrounds (CZE=Czech, ENG=English, NOR=Norwegian) in the two speaking styles.

To describe the internal temporal structure of word realisations, we calculated two ratios: the proportion of the duration of the vowel within the whole item duration, and the proportion of the phonetically voiced portion within the fricative, both expressed as percentages. Figure 3.20 shows the mean values of both of these measures (proportion of vowel within the item and proportion of voicing within the fricative) for groups based on the speakers' L1 backgrounds and speaking style. As to the proportion of

vowel within the item duration, Norwegians had the vowel proportion higher (54%) than the other two speaker groups (45% for Czech speakers and 47% for native English speakers). The vowel proportions for the two speaking styles were 45% for read speech and 52% for spontaneous speech.

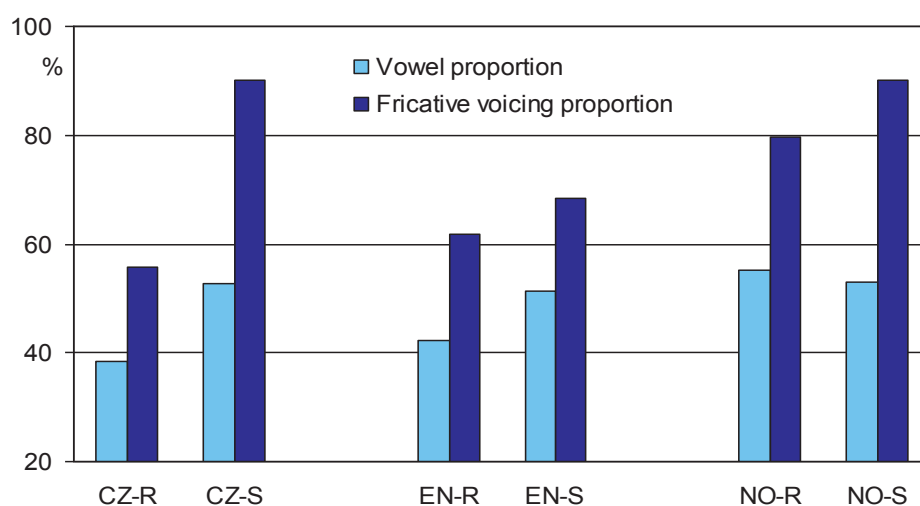


Figure 3.20: Means of vowel proportions within the item and proportions of voicing within the fricative (in %) in the word *of*, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

An analysis of variance showed significant effects of both the speakers' L1 background ($F(2, 288) = 7.31$; $p = 0.001$) and speaking style ($F(1, 288) = 13.4$; $p < 0.001$) as well as a significant interaction between these two factors ($F(2, 288) = 6.23$; $p = 0.002$). Tukey's post-hoc test showed that the effect of the speakers' L1 background was due to the Norwegians' items being significantly different from both the Czech ($p = 0.001$) and the native English productions ($p = 0.006$). Also, the interaction of L1 background and speaking style factors was due to the fact that in Norwegian productions the vowel proportion did not differ too much between read speech (55%) and spontaneous speech (53%), while in the other two speaker groups, the differences were much larger, showing noticeably greater ratio values in spontaneous speech (for Czech speakers 38%

Chapter 3

and 53%, for English speakers 42% and 51%, for read and spontaneous speech respectively).

The proportion of the phonetically voiced fricative part within the duration of fricative also differed in the three speaker groups, being 72% for the Czech speakers, 65% for the native English speakers and 85% for the Norwegian speakers. There was also a difference between the two speaking styles: in read items the voicing ratio was 65% while in spontaneous speech it was 83%. Levene's test also indicated unequal variances ($p < 0.001$) in this case, and therefore, we carried out non-parametric analyses. Kruskal-Wallis tests revealed significant effects of both L1 ($\chi^2(2) = 16.5$; $p < 0.001$) and speaking style ($\chi^2(1) = 18.4$; $p < 0.001$). A post-hoc test using Mann-Whitney tests with Bonferroni correction showed that the L1 effect was again due to the Norwegians' items being significantly different from both Czech ($p = 0.027$) and native English productions ($p < 0.001$). In addition, we could observe an interaction of L1 background with speaking style. Even though the voicing proportions were consistently higher in spontaneous speech for all three speaker groups, the size of differences between styles in the three speaker groups differed, causing the interaction of L1 background and speaking style. While for Norwegians, the ratios for read and spontaneous speech amounted to 80% and 90%, and for English speakers 62% and 68%, respectively, the ratios in the Czech speaker group differed much more, with 56% in read speech and 90% in spontaneous speech. This result can be explained by the somewhat long fricative durations in read speech in the Czech speaker group (read vs. spontaneous: 89 ms vs. 53 ms). The English and Norwegian speakers had similar fricative durations for these two conditions (52 ms vs. 50 ms for English speakers and 50 ms vs. 49 ms, for read and spontaneous speech, respectively). In order to test this interaction, we performed Kruskal-Wallis tests separately for each speaker group. These tests revealed that while in the native speaker group there was no significant effect of speaking style ($\chi^2(1) < 1$), the speaking style effect was significant for Norwegian speakers ($\chi^2(1) = 4.49$; $p = 0.034$) and highly significant for the Czech speaker group ($\chi^2(2) = 21.9$; $p < 0.001$).

3.3.3.2 Spectral measures in the word of

As a complement to the durational measurements we used two Bark distance measures, F1 - F0 and F3 - F2, as measures of vowel quality. These two measures were found to correspond to vowel openness and vowel backness, respectively (Syrdal and Gopal 1986; see Section 3.2.3.5 for more details). Mean F1 - F0 and F3 - F2 Bark distance values for speaker groups with different L1 backgrounds and different speaking styles are presented in Figure 3.21.

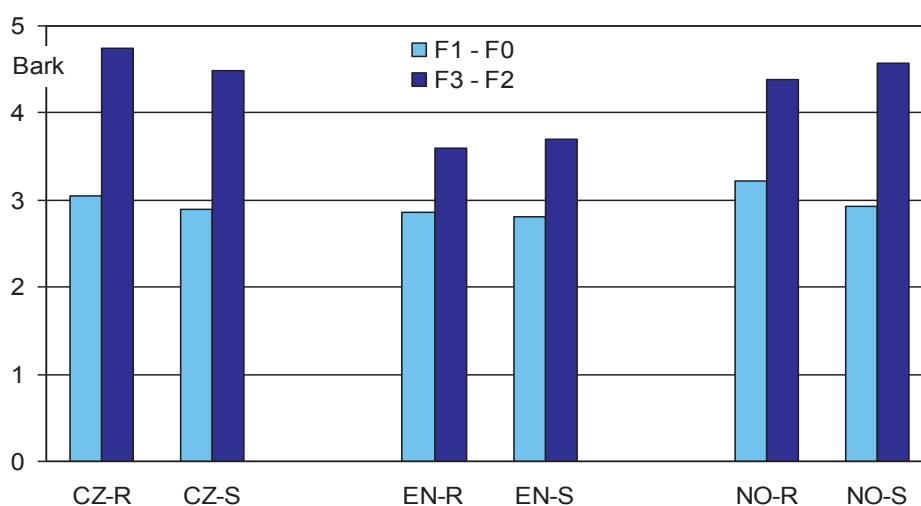


Figure 3.21: Mean Bark distances of formants (F1 - F0, F3 - F2) in the vowel of the word *of*, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

In our data, we found a slight difference in the values of F1 - F0 between the two speaking styles (3.0 Bark in read speech and 2.9 Bark in spontaneous speech). Lower values in spontaneous speech would correspond to less open vowels. The mean values were 3.0 Bark for the Czech speakers, 2.8 Bark for the native English speakers and 3.1 Bark for the Norwegians. An analysis of variance showed that the effect of speaking style was significant ($F(1, 265) = 3.96$; $p = 0.048$), while the effect of speakers' L1 background reached only marginal significance ($F(2, 265) = 3.00$; $p = 0.051$). There was

Chapter 3

no significant interaction between these two factors. In contrast to the F1 - F0 Bark distance, for the F3 - F2 measure a highly significant effect of L1 background was found ($F(2, 287) = 29.8$; $p < 0.001$). Neither speaking style (both read and spontaneous speech: 4.2 Bark) nor its interaction with the factor L1 background, however, reached statistical significance ($F(1, 287) < 1$ and $F(2, 287) = 1.44$; $p = 0.240$, respectively). Tukey's post-hoc test showed that that F3 - F2 values for the Czech (4.6 Bark) as well as the Norwegian speakers (4.5 Bark) were significantly different from the native English values (3.6 Bark); ($p < 0.001$, for both). This indicates more peripheral vowel qualities for non-native speakers.

Although we assumed that the Bark distance measures eliminated the gender-specific differences in the formant values, we considered it appropriate to test this on our material. Therefore, we compared males' and females' means of both formant distances. When we inspected the realisations of the word *of*, it seemed that the F3 - F2 values of males and females differed somewhat (means being 4.1 vs. 4.4 Bark for females and males, respectively). An analysis of variance with L1 background, speaking style and gender as factors, however, showed, apart from the significant effect of L1 background, a strong interaction of L1 background and gender ($F(2, 281) = 6.73$; $p = 0.001$). The gender effect was even more noticeable in the F1 - F0 Bark distance measure. The mean values for females and males were 2.7 and 3.3 Bark, respectively. An analysis of variance with L1 background, speaking style and gender as factors showed only a significant effect of gender ($F(1, 259) = 78.6$; $p < 0.001$). The results of corresponding comparisons inspecting the realisations of the word *to* showed very similar tendencies (namely a large effect of gender on F1 - F0 Bark distance and an interaction of L1 background effect and gender on F3 - F2 Bark distance). These findings highlighted the need to analyse the data more thoroughly to ensure that the gender differences do not distort the statistical results of other factors' effects. To obtain more detailed information we analysed females' and males' formant values separately, focussing on F1 and F2.

An analysis of variance with the factors L1 background and speaking style, inspecting the first formant values in the preposition *of* produced by females found no significant effects. A similar analysis of the items produced by male speakers found only a significant interaction of the effects of L1 background and speaking style ($F(2, 131)=3.11$; $p=0.048$). This interaction was probably due to the Norwegian male speakers' large difference between F1 values in read (4.5 Bark) and spontaneous (4.1 Bark) speech, compared to smaller differences in both the native English and Czech speaker groups. Similarly, we inspected the values of the the second formant. The means and standard deviations of F2 values for females and males with different L1 backgrounds are listed in Table 3.6.

Table 3.6: Mean F2 values (in Bark) and F2 standard deviations (in Bark) in the function word *of*, for females and males with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian).

		mean (Bark)	std. dev. (Bark)
CZ	females	10.1	0.95
	males	9.0	0.82
EN	females	11.7	0.95
	males	10.5	0.57
NO	females	10.5	0.95
	males	9.7	0.90

An analysis of variance with the factors L1 background and speaking style found that for females the L1 background effect was highly significant ($F(2, 151)=41.7$; $p<0.001$). Tukey's post-hoc test showed that native speakers' values differed significantly from the values of both non-native groups. An analysis of variance with the factors L1 background and speaking style inspecting the items produced by males also found a significant effect of L1 background ($F(2, 131)=33.0$; $p<0.001$). Tukey's post-hoc test showed that the native males had significantly higher F2 values than both non-native male groups, and the Czech males' values were still significantly lower than the values of Norwegian males. These results are in good correspondence with the Bark distance results, confirming no effects of L1 background on the degree of vowel openness, and a strong effect of L1 background on the front-back dimension of vowel articulation in the

Chapter 3

realisations of the function word *of*. Both the data from the females and males showed that natives produced significantly more fronted vowels than both non-native groups.

To inspect the spectral quality of fricatives, we used the friction intensity measure of the difference between the high-frequency band (over 5000 Hz) and the low-frequency band (below 5000 Hz) energy in dB (more details can be found in Section 3.2.3.5). This measure is intended to describe the amount of high-frequency friction noise present in the spectrum, and it should correspond to the fortis character of the fricatives. Since the values of friction intensity strongly reflect the amount of voicing in the measured segment, the friction intensity measure is highly inversely correlated with the proportion of fricative voicing ($r = -0.827$; $N = 280$; $p < 0.001$), and we can therefore expect considerable similarities between the voicing proportion and friction intensity results. On the other hand, friction intensity represents an independent measure of spectral properties, not reliant on the manual annotation of phonetic voicing, and thus it can be seen as a useful complement to the manually determined voicing proportion measure. The mean friction intensity values were somewhat lower for the Norwegian speaker group (-34 dB), than for the other speaker groups (Czech speakers: -24 dB, English speakers: -19 dB). Also the mean value for read speech (-22 dB) differed from the mean for spontaneous items (-29 dB). As in the measure of the voicing proportion in the fricative, Levene's test indicated significantly unequal variances ($p < 0.001$). Kruskal-Wallis tests confirmed significant effects of both L1 background ($\chi^2(2) = 41.9$; $p < 0.001$) and speaking style ($\chi^2(1) = 11.4$; $p < 0.001$). Moreover, the Kruskal-Wallis tests carried out for each speaker group separately showed that speaking style effect was only significant for the Czech speaker group ($\chi^2(1) = 11.1$; $p < 0.001$) while for the natives and Norwegian speakers, it did not reach significance ($\chi^2(1) = 1.28$; $p = 0.257$ and $\chi^2(1) < 1$, respectively). As expected, these results are very similar to the results relating to fricative voicing proportion.

3.3.3.3 Influence of context on the realisations of the word of

In this section we will inspect the influence of some aspects of context on various parameters in the production of the preposition *of*. This includes the influence of the type (vowel or consonant) of neighbouring segments, and the influence of phonetic voicing in the neighbouring segments on relevant measures in the realisations of the preposition *of*. With regard to the previous research (cf. Section 1.2.1.1) it seemed relevant to inspect the influence of context type on measures relating to temporal organisation. In addition, we observed the context type influence on the first formant measure (Bark distance F1 - F0) indicating the degree of openness of the vowel. As for the influence of phonetic voicing, we inspected the temporal measures, and the measure of friction intensity in the fricative.

The normalised durations seemed to differ somewhat depending on the type of segment both preceding and following the preposition *of*. The mean normalised durations were 107 ms for words preceded by consonants and 117 ms for items preceded by vowels. The mean normalised word duration values for items depending on the immediately following segment type were 105 ms and 116 ms, for words followed by consonants and vowels, respectively. Figure 3.22 displays the means of normalised durations in the word *of* followed by different segment types (consonant or vowel) for different speaker groups and speaking styles.

We can observe that apart from the group of spontaneous items produced by the Czech speakers, the normalised durations are longer in items followed by vowels. Since Levene's test indicated unequal variances with high significance ($p < 0.001$) we performed the statistical tests using non-parametric tests inspecting the effect of left and right context type. Kruskal-Wallis tests showed that while there was no significant effect of the preceding segment type ($\chi^2(1) = 2.16$; $p = 0.141$), the following segment type affects the normalised durations significantly ($\chi^2(1) = 4.91$; $p = 0.027$).

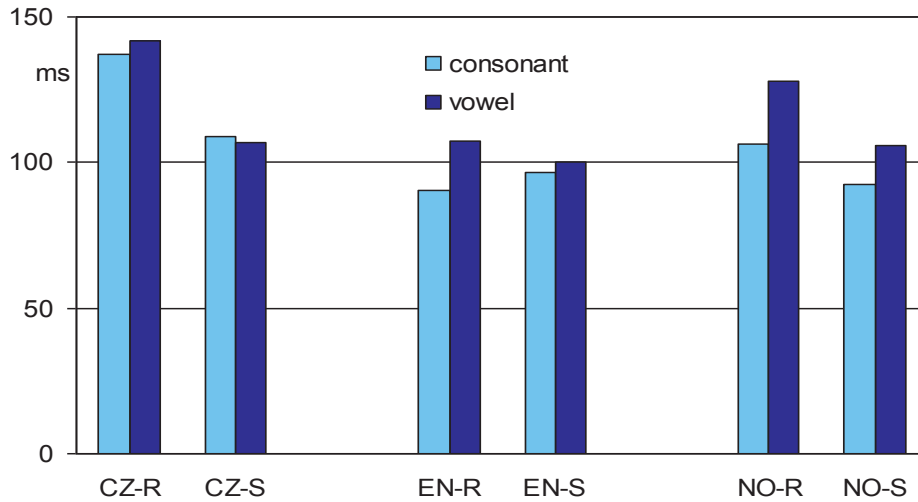


Figure 3.22: Normalised duration means (in ms) in the word *of* followed by different segment types (consonant or vowel), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

The vowel proportion within the word duration seemed to be influenced by the right segment type as well, the mean vowel proportions being 50% for items followed by consonants and 43% for items followed by vowels. On the contrary, the preceding segment type had almost no effect (means being 49% vs. 51% for items preceded by consonants and vowels, respectively). An analysis of variance with the factors L1 background, speaking style, left segment type and right segment type, however, showed neither a significant effect of L1 background and speaking style, nor a significant effect of the preceding or following segment type on this variable.

The proportion of phonetically voiced fricative part within the duration of fricative also proved to be slightly influenced by both the preceding and following segment types. The mean fricative voicing proportions were 74% for items preceded by consonants and 69% for items preceded by vowels. The means for items depending on the type of following segment were 72% (for items followed by consonants) and 79% (for items followed by vowels). Here, too, Levene's test indicated unequal variances with high significance ($p < 0.001$). Therefore, we performed the statistical tests using non-

parametric tests inspecting the effect of left and right context type. Kruskal-Wallis tests showed that neither the preceding segment type ($\chi^2(1) < 1$) nor the following segment type ($\chi^2(1) = 2.44$; $p = 0.118$) had a significant effect on the proportion of voicing in fricatives.

As for the measures of vowel quality, we found that the mean F1 - F0 Bark distance, corresponding to vowel openness, varied consistently depending on the type of preceding segment. The values were lower when the preposition *of* followed consonants (2.9 Bark) than when it followed vowels (3.4 Bark). This finding is not surprising, considering the expected positions of the articulators for the basic classes of speech sounds, specifically the presumably more open vocal tract for the articulation of vowels. The F1 - F0 values for items followed by segments of different types did not differ at all (3.0 Bark for items followed by both types of speech sounds). The mean F1 - F0 values for the three speaker groups and different speaking styles, depending on the type of preceding segment are presented in Figure 3.23.

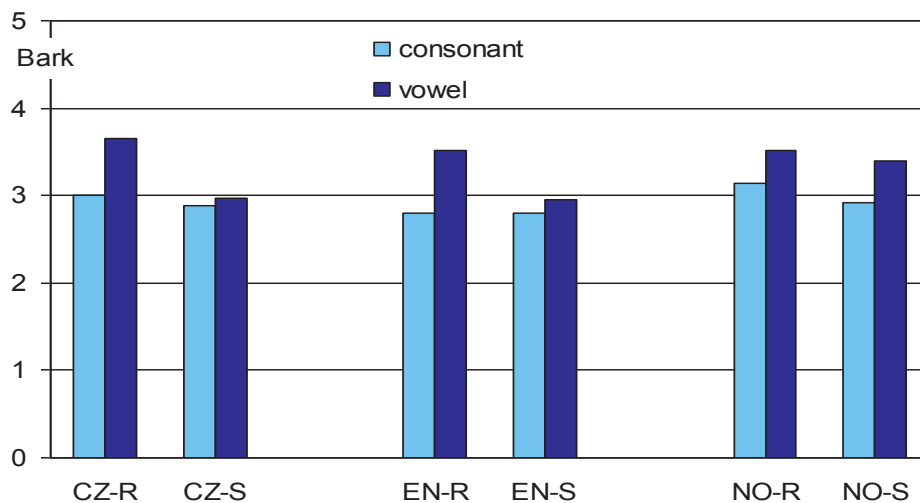


Figure 3.23: Mean F1 - F0 distances (in Bark) in the vowel of the word *of* preceded by different segment types (consonant or vowel), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

Chapter 3

It can be observed that the effect of the preceding segment type is consistent for all speaker groups with different L1 backgrounds and both speaking styles. An analysis of variance with the factors L1 background, speaking style, left segment type and right segment type, however, showed only a significant effect of speaking style ($F(1, 248)=4.77$; $p=0.030$) but surprisingly no significant effect of preceding segment type ($F(1, 248)=3.29$; $p=0.071$). Since previous analyses have shown that the F1 - F0 Bark distance values in our data differ between speakers of different genders, we also carried out an analysis of variance with gender as a factor, apart from the previously mentioned factors. The results of this analysis confirmed a significant effect of preceding segment type ($F(1, 231)=3.99$; $p=0.047$) and an effect of speaker gender ($F(1, 231)=5.72$; $p=0.018$) on the F1 - F0 Bark distance measure.

In addition to analysing the effects of the type of neighbouring segments (consonant or vowel), we examined the influence of phonetic voicing in the neighbouring segments on the realisation of the function word *of*. The normalised durations were considerably shorter in items preceded by a voiceless segment (101 ms) as compared to items preceded by a voiced segment (113 ms). The following segment voicing, on the other hand, did not seem to have an effect on the normalised duration of items (108 vs. 107 ms, for items followed by voiceless or voiced segment, respectively). Since Levene's test indicated unequal variances with high significance ($p < 0.001$) we performed the statistical tests using non-parametric tests inspecting the effect of phonetic voicing in preceding and following segments. Kruskal-Wallis tests showed that there was a significant effect of preceding segment voicing ($\chi^2(1)=8.85$; $p=0.003$), but as could be expected, the effect of phonetic voicing in the following segment did not reach significance ($\chi^2(1) < 1$).

The proportion of vowel within the word duration also seemed to be influenced by voicing in the neighbouring segments. The mean vowel proportions were 46% when preceded by a voiceless segment, and 51% when preceded by a voiced segment. The proportion also varied depending on the phonetic voicing in the segment following the preposition, amounting to 46% vs. 50% for items followed by a voiceless and voiced

segment, respectively. An analysis of variance with L1 background, speaking style and presence of phonetic voicing in the preceding and following segment as factors, showed, apart from significant effects of the speakers' L1 background and speaking style, and an interaction between these two factors, a significant effect of phonetic voicing in the following segment as well ($F(1, 270) = 4.71$; $p = 0.031$). The effect of phonetic voicing in the preceding segment, however, did not reach significance.

When we consider the phonetic voicing in neighbouring segments, an especially large effect could be expected on the proportion of voicing within the fricative. While voicing in the preceding segment did not have a large influence on the voicing proportion (72% vs. 76% for items preceded by voiceless and voiced segments, respectively), the voicing proportion difference due to voicing in the following segment was considerable (52% vs. 85% for items followed by voiceless and voiced segments, respectively). Figure 3.24 shows the influence of following segment voicing on the proportion of fricative voicing for speaker groups with different L1 backgrounds and different speaking styles.

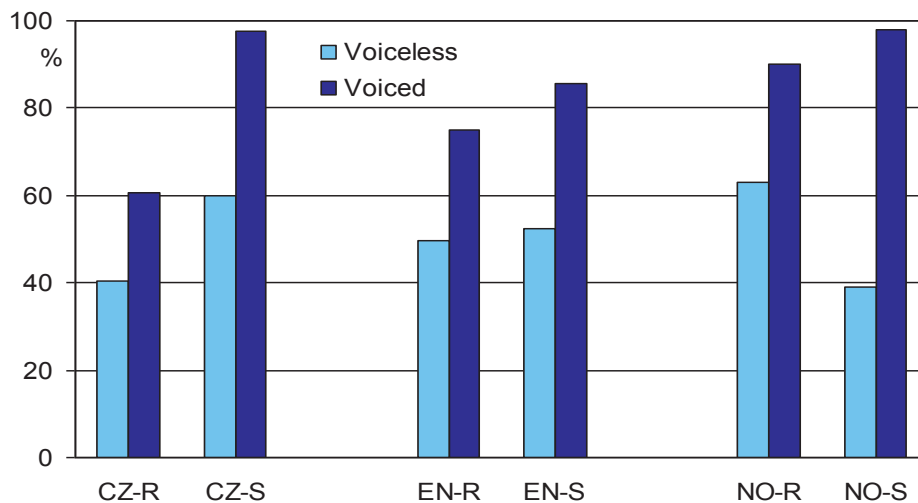


Figure 3.24: Proportions of voicing in the fricative (in %) in the word *of* followed by a voiceless/voiced segment, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

Chapter 3

Here Levene's test also indicated unequal variances with high significance ($p < 0.001$). Therefore, we performed the statistical tests using non-parametric tests inspecting the effect of voicing in the preceding and following segment. Kruskal-Wallis tests showed that while there was no significant effect of preceding segment voicing ($\chi^2(1) < 1$), the voicing in the following segment had a highly significant effect on the proportion of voicing within the fricative ($\chi^2(1) = 62.6$; $p < 0.001$). Moreover, we can observe larger differences between items with voiceless and voiced context in spontaneous speaking style (52% vs. 95%) than in read speech (52% vs. 74%), as can be observed in Figure 3.24. The Kruskal-Wallis tests carried out separately for read and spontaneous items showed that the right context voicing effect was highly significant both in read ($\chi^2(1) = 11.5$; $p < 0.001$) and spontaneous speech ($\chi^2(1) = 70.3$; $p < 0.001$).

Lastly, we inspected the influence of context voicing on friction intensity, an acoustic variable closely related to the proportion of voicing in the fricative. As was expected, we found a large difference in the mean values due to voicing in the following segment (-15 dB vs. -31 dB for items with voiceless vs. voiced right context). As with the previous measure of fricative voicing proportion, here, too, Levene's test indicated unequal variances with high significance ($p < 0.001$). Therefore, we performed the statistical tests using non-parametric tests inspecting the effect of phonetic voicing in the preceding and following segment. Kruskal-Wallis tests showed that while there was no significant effect of preceding segment voicing ($\chi^2(1) < 1$), the voicing in the following segment had a highly significant effect on the friction intensity in the fricative ($\chi^2(1) = 58.3$; $p < 0.001$).

Apart from describing how context factors may influence the realisations of the preposition *of*, we were also interested in the distribution of the various aspects of context in the present dataset. First we inspected the distributions of segment types of the immediately neighbouring left and right context, comparing the groups of items based on the speakers' L1 background and speaking style (see Tables 3.4 and 3.5 on page 98). Pearson's Chi-squared tests showed that neither the left segment type distributions nor the right segment type distributions varied significantly between the

groups ($\chi^2(5)= 8.43$; $p= 0.134$ and $\chi^2(5)= 7.42$; $p= 0.191$, for left and right segment type distributions, respectively). In a similar way, we inspected the distributions of voiced/voiceless segments preceding and following the word *of*, for the subgroups based on the speakers' L1 background and speaking style. Pearson's Chi-squared tests showed that while the distributions of phonetically voiced/voiceless preceding segments did not vary significantly between the groups ($\chi^2(5)= 4.75$; $p= 0.447$), the distributions of phonetically voiced/voiceless segments following the tokens of the function word *of* differed significantly between the groups ($\chi^2(5)= 26.5$; $p< 0.001$). The distributions of phonetically voiced/voiceless segments following the word *of* for groups of items based on the speakers' L1 background and speaking style are displayed in Figure 3.25. We can observe that the number of tokens with a voiceless following segment is noticeably higher among the English items in both speaking styles. The possibility of confounding the effects of L1 background and speaking style with the effects of unevenly distributed voiced/voiceless following segments has to be taken into consideration when interpreting the results.

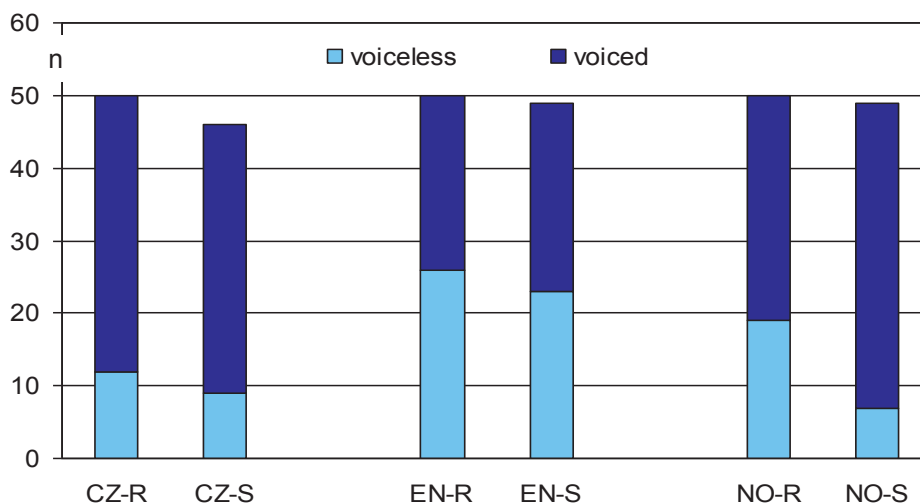


Figure 3.25: Distributions of phonetically voiced/voiceless segments following the word *of* (numbers of tokens), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

3.3.4 TO

3.3.4.1 Temporal organisation in the word *to*

As in the previous sections, we first inspected the normalised durations of the items. Figure 3.26 shows boxplots of the normalised durations of the word *to* in read and spontaneous speech for the three speaker groups based on the L1 background. The boxes in the plots contain the inter-quartile range, and the whiskers extend to the furthest datapoint that lies within 1.5 times the inter-quartile range from the box on each side.

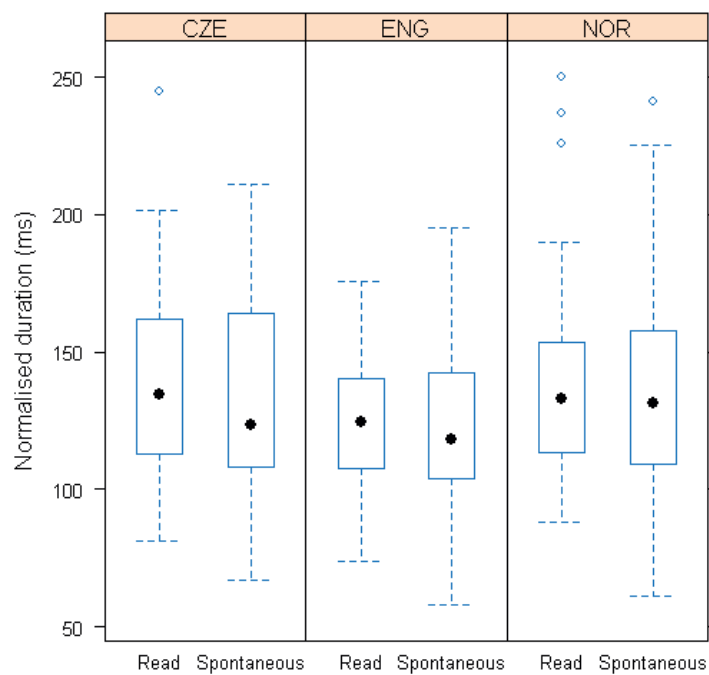


Figure 3.26: Normalised word durations (in ms) in the word *to*, for speaker groups with different L1 backgrounds (CZE=Czech, ENG=English, NOR=Norwegian) in the two speaking styles.

As in the previous sections, the normalised durations of this function word differed between the speaker groups with different L1 backgrounds. The mean normalised durations were 136 ms for the Czech speakers, 124 ms for the native English speakers and 137 ms for the Norwegian speakers. On the other hand, the mean durations did not differ much between the speaking styles (133 ms in read speech vs. 132 ms in spontaneous speech). An analysis of variance showed a significant effect of L1 background ($F(2, 279) = 4.79$; $p = 0.009$), but no significant effect of speaking style, nor a significant interaction between the two factors ($F(1, 279) < 1$ and $F(2, 279) < 1$, respectively). Tukey's post-hoc test showed that while the English speakers' normalised durations differed significantly from both the Czech speakers ($p = 0.030$) and the Norwegian speakers ($p = 0.014$), the difference between the two non-native groups was not significant ($p = 0.969$).

As in the previously investigated function words, we calculated the ratios using durations of segments and segment phases: the proportion of the vowel within the duration of the word, and the proportion of the release duration within the duration of plosive, both expressed as percentages. Figure 3.27 shows the mean values of both of these measures (vowel proportion within the item and proportion of release phase within the plosive) for groups based on the speakers' L1 background and speaking style.

As to the vowel proportion, the mean values differed between the speaker groups with different L1 backgrounds, English vowel proportion being lower (21%) than the proportion in the two non-native groups (30% for Czech speakers and 28% for Norwegians). The percentages for the two speaking styles were 26% for read speech and 27% for spontaneous speech. An analysis of variance showed only a significant effect of L1 background ($F(2, 278) = 17.7$; $p < 0.001$). Tukey's post-hoc test showed that the L1 background effect was due to the native English speakers' items being significantly different from both the Czech speakers' ($p < 0.001$) and the Norwegians' productions ($p < 0.001$). The difference between the two non-native groups turned out to be non-significant ($p = 0.288$).

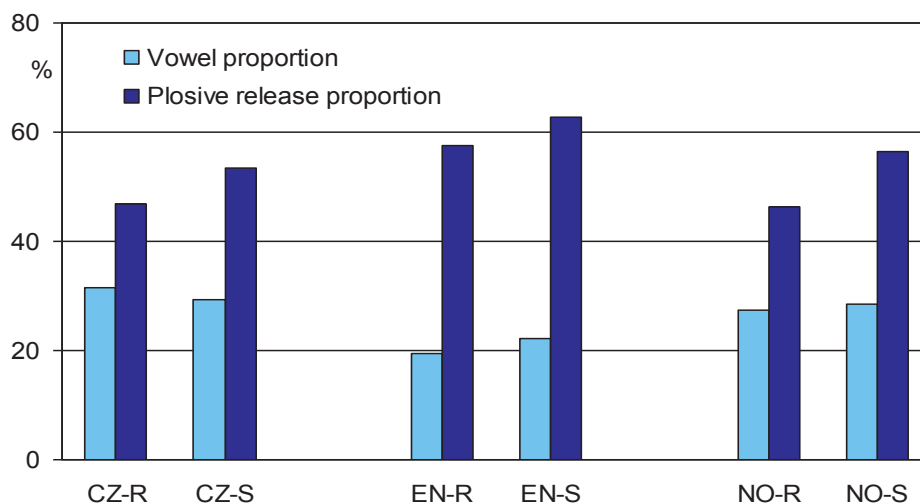


Figure 3.27: Means of vowel proportions within the item and proportions of release within the plosive (in %) in the word *to*, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

For all three speaker groups the release of the plosive closure constitutes a considerable part of the total consonant duration (pooled across all conditions: 54%). The mean values varied between the speaker groups (Czech speakers: 50%; native English speakers: 60%; Norwegians: 51%) as well as between read speech (50%) and spontaneous speech (58%). According to an analysis of variance, the influence of both L1 background and speaking style was highly significant ($F(2, 278) = 15.2$; $p < 0.001$ and $F(1, 278) = 21.2$, $p < 0.001$) with no significant interaction ($F(2, 278) < 1$). Tukey's post-hoc test showed that the effect of L1 background was due to the relative release durations being significantly smaller for both groups of non-native speakers than for the natives (in both cases $p < 0.001$). Here as well the difference between the two non-native groups was not significant ($p = 0.799$). In connection with this measure of proportion of plosive release and closure, we also inspected the distribution of cases where a complete plosive closure was not present (i.e. was realised as a constriction). The total number of such cases was 33 out of the total of 285 items, and they were distributed quite evenly among the speaker groups (13 tokens for the Czech speaker group, 11 tokens for the native English speakers and 9 tokens for the Norwegian speaker group; $\chi^2(2) < 1$). We

could, however, observe that this phenomenon is significantly more frequent in spontaneous speech (23 out of the total 33 tokens; $\chi^2(1)= 5.12$; $p= 0.024$). This fact may indicate a tendency to a decreased articulatory effort in the production of plosives in spontaneous speech, which is also in agreement with the higher proportion of release phase (i.e. lower durational proportion of closure) in spontaneous speech. The total number of cases without complete closure was not sufficient for more detailed analyses.

3.3.4.2 Spectral measures in the word *to*

As with the preposition *of*, we measured the formant values in the vowel of the word *to* and determined the Bark distance values F1 - F0 and F3 - F2, corresponding to vowel openness and vowel backness. The F1 - F0 and F3 - F2 values for the three speaker groups with different L1 backgrounds and different speaking styles are presented in Figure 3.28.

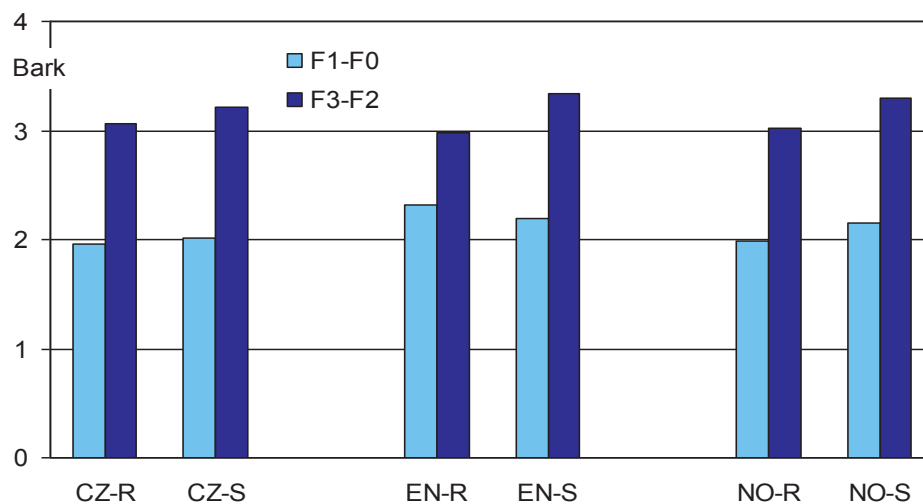


Figure 3.28: Mean Bark distances of formants (F1 - F0, F3 - F2) in the vowel of the word *to*, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

Chapter 3

The values of the first formant measure F1 - F0 did not vary much between the speaker groups with different L1 backgrounds (Czech speakers: 2.0 Bark; native English speakers: 2.3 Bark; Norwegians: 2.1 Bark), nor between the speaking styles (both read and spontaneous speaking style: 2.1 Bark). An analysis of variance showed neither a significant effect of L1 background and speaking style, nor a significant interaction between the two factors ($F(2, 226) = 2.60$; $p = 0.076$, $F(1, 226) < 1$ and $F(2, 226) < 1$, respectively).

The means of the F3 - F2 values in the word *to* were 3.1 Bark for all three speaker groups, and there was only a slight difference between the speaking styles, with 3.0 Bark in read speech and 3.3 Bark in spontaneous speech. An analysis of variance found only a significant effect of speaking style ($F(1, 254) = 7.27$; $p = 0.007$) but neither any effect of L1 background, nor its interaction with speaking style (in both cases $F(2, 254) < 1$).

Since previous analyses revealed that the Bark distance measures retain gender-related variability in our material (see Section 3.3.3.2), we carried out additional analyses inspecting the raw values of the first and second formant for female and male speakers separately. The mean values and standard deviations of the F1 values in Bark for females and males with different L1 backgrounds are listed in Table 3.7.

Table 3.7: Mean F1 values (in Bark) and F1 standard deviations (in Bark) in the function word *to*, for females and males with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian).

		mean (Bark)	std. dev. (Bark)
CZ	females	3.8	0.42
	males	3.3	0.31
EN	females	4.3	0.96
	males	4.0	0.54
NO	females	3.7	0.46
	males	3.4	0.46

We can observe that the F1 values in Bark differed considerably between the speaker groups with different L1 backgrounds, both for females and males. The mean F1 values of the native English speakers were noticeably higher than those of the non-native speakers, indicating more open vowels. Moreover, we can observe higher standard deviations in the native English speaker group, especially for females (0.96 Bark). This indicates a higher variability of the F1 values produced by the native speakers as compared to those in the non-native productions. The statistical analysis of the female data had to be carried out using non-parametric tests, since Levene's test indicated unequal variances at a high level of significance ($p < 0.001$). Kruskal-Wallis tests revealed a significant effect of L1 background ($\chi^2(2) = 17.1$; $p < 0.001$) but no effect of speaking style ($\chi^2(1) < 1$) in the female data. A post-hoc test using Mann-Whitney tests with Bonferroni correction showed that the effect of L1 background was due to the native English females' values being significantly different from both the Czech ($p = 0.009$) and Norwegian females' productions ($p < 0.001$). As for the male speakers, an analysis of variance with the factors L1 background and speaking style revealed a highly significant effect of L1 background ($F(2, 118) = 22.1$; $p < 0.001$) but no significant effect of speaking style or interaction ($F(1, 118) = 3.59$; $p = 0.061$ and ($F(2, 118) < 1$, respectively). Tukey's post-hoc test showed that similarly as in females, the effect of L1 background was due to the significantly higher F1 values of native male speakers compared to both groups of non-native male speakers (in both cases $p < 0.001$). The difference between the Czech and Norwegian male speakers' values was not significant ($p = 0.946$). In sum, these analyses confirmed that there were significant differences in the degree of vowel openness in productions of the function word *to* between the native English speakers (formant values indicating more open vowels) and both non-native groups. These were not revealed by an analysis of F1 - F0 Bark distances pooled across female and male data, presumably due to the remaining gender-related variation in Bark distance values.

Similarly, we inspected values of the second formant. The mean values and standard deviations of the second formant values in Bark for females and males with different L1 backgrounds are listed in Table 3.8.

Chapter 3

Table 3.8: Mean F2 values (in Bark) and F2 standard deviations (in Bark) in the function word *to*, for females and males with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian).

		mean (Bark)	std. dev. (Bark)
CZ	females	11.7	1.13
	males	10.6	0.71
EN	females	12.4	0.75
	males	10.9	0.69
NO	females	11.6	0.77
	males	11.1	0.69

An analysis of variance with the factors L1 background and speaking style found that for females, the effect of L1 background was highly significant ($F(2, 129) = 11.0$; $p < 0.001$) but neither speaking style nor the interaction of L1 background with speaking style reached significance ($F(1, 129) = 1.51$; $p = 0.221$ and $F(2, 129) < 1$). Tukey's post-hoc test showed that the native English female speakers' values differed significantly both from the Czech female speaker group ($p = 0.001$) and from the Norwegian female speaker group ($p < 0.001$). For the items produced by males, an analysis of variance with the factors L1 background and speaking style also found a significant effect of L1 background ($F(2, 119) = 4.98$; $p = 0.008$) as well as a significant effect of speaking style ($F(1, 119) = 3.98$; $p = 0.048$). This was due to the second formant values being lower in spontaneous speech (10.7 Bark) than in read speech (11.0 Bark). There was no significant interaction of L1 background and speaking style ($F(2, 119) < 1$). Tukey's post-hoc test showed that there was only a significant difference between the Norwegian and Czech male speaker groups ($p = 0.005$). The native English male speakers did not differ significantly from either the Czech speakers ($p = 0.234$) or the Norwegians ($p = 0.647$). The combined results of the F2 analyses indicate that there may be differences related to the L1 background, but the females and males in our data did not show consistent tendencies. For the females, the native English speakers had higher F2 values (corresponding to more fronted vowels) than the non-natives, while for the males the Norwegian group had the highest F2 values. The effect of speaking style was only significant in the data from males, although a closer analysis of the data confirmed a similar tendency in the female group (pooled across the speaker groups with different

L1 backgrounds, raw F2 values for females were 12.05 Bark in read speech vs. 11.85 Bark in spontaneous speech).

As for the fricatives in the preposition *of*, we used a measure of friction intensity in the plosive release phase in the function word *to* to determine the spectral quality of the plosives. Inspection of the mean values showed somewhat lower values for the Norwegian speaker group (-11 dB), than for the other groups (Czech speakers: -6 dB, English speakers: -5 dB). The mean value for read speech (-8 dB) also slightly differed from the mean from spontaneous items (-6 dB). An analysis of variance confirmed an effect of L1 background and speaking style, but found no significant interaction of the two factors ($F(2, 278) = 24.6$; $p < 0.001$, $F(1, 278) = 6.17$; $p = 0.014$ and $F(2, 278) = 1.37$; $p = 0.257$). Tukey's post-hoc test showed that the effect of L1 background was due to friction intensity values being lower in the Norwegian speaker group, than in both other groups (in both cases $p < 0.001$), while the difference between the English and Czech speakers' friction intensity values was not significant ($p = 0.352$).

3.3.4.3 Influence of context on the realisations of the word *to*

As in Section 3.3.3.3, this section will examine the influence of neighbouring segment type and voicing on selected parameters of the realisation of the function word *to*. We considered it relevant to inspect the influence of segment type on measures relating to temporal organisation and on the first formant measure indicating the degree of openness of the vowel. As for the influence of phonetic voicing in the neighbouring segment, we inspected the normalised duration measure, vowel proportions, and the measure of friction intensity in the plosive release.

The item durations varied considerably depending on the type of following segment, the mean normalised durations being 130 ms for items followed by consonants and 157 ms for items followed by vowels. Figure 3.29 displays the normalised duration means in the word *to* for speaker groups with different L1 backgrounds and different speaking styles, depending on the type of following segment. It can be seen that the items followed by

Chapter 3

vowels have consistently longer normalised durations than items followed by consonants. The type of preceding segment, on the other hand, did not have a very noticeable effect (mean values being 131 ms vs. 136 ms for items preceded by consonants and vowels, respectively). An analysis of variance with the factors L1 background, speaking style, left segment type and right segment type showed, as expected, only a significant effect of right segment type ($F(2, 263) = 6.66$; $p = 0.010$). No other effect or interaction reached significance.

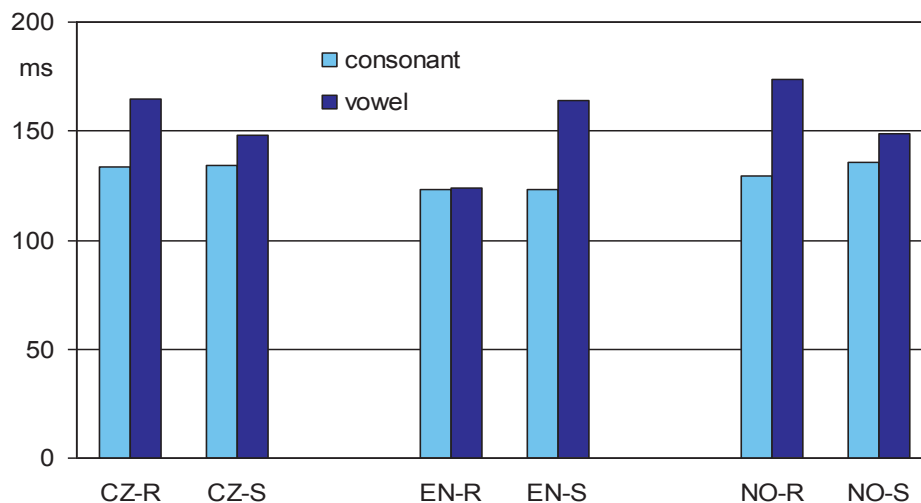


Figure 3.29: Normalised duration means (in ms) in the word *to* followed by different segment types (consonant or vowel), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

The vowel proportions also varied depending on the type of following segment (mean vowel proportions being 26% vs. 30% for items followed by consonants and vowels, respectively). An analysis of variance with the factors L1 background, speaking style, and left and right segment type showed only a marginally significant effect of right segment type ($F(1, 262) = 3.80$; $p = 0.052$), while no other effect or interaction reached significance. The proportion of plosive release, on the other hand, seemed to be influenced by the preceding segment type. The mean release proportions were higher for words preceded by consonants (55%) than for words preceded by vowels (49%). It

is important to realise that the release proportion is a percentual complement to closure proportion, and the higher values of release proportion thus indicate a shortened closure phase. Surprisingly, an analysis of variance with the factors L1 background, speaking style, and left and right segment type showed only a significant effect of speaking style ($F(1, 262) = 6.27$; $p = 0.013$) but no significant effect of left segment type ($F(1, 262) = 1.44$; $p = 0.232$), nor any other effects or interactions.

As for the influence of neighbouring segment type on vowel openness, the values of the F1 - F0 Bark distance varied depending on the type of following segment. Mean F1 - F0 Bark distances depending on the following segment type for speaker groups with different L1 backgrounds in the two speaking styles, are shown in Figure 3.30.

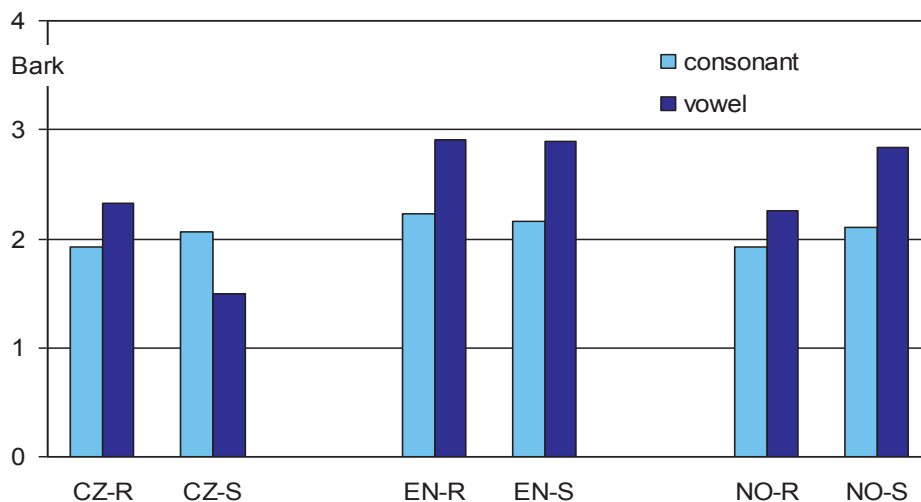


Figure 3.30: Mean F1 - F0 distances (in Bark) in the vowel of the word *to* followed by different segment types (consonant or vowel), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

The Bark distance values were lower when the preposition *to* was followed by consonants (2.0 Bark) than when it was followed by vowels (2.4 Bark). Similar to the effect of left segment type on the first formant measure in the word *of*, this difference is likely to be a consequence of the degree of opening of the articulators in the speech

Chapter 3

sound directly adjacent to the observed vowel. An analysis of variance with the factors L1 background, speaking style, and left and right segment type showed only a significant effect of right segment type ($F(1, 210) = 6.40$; $p = 0.012$). Since previous analyses have shown that the $F1 - F0$ Bark distance differs depending on speakers' gender, we also carried out an analysis of variance with gender as a factor, apart from the previously mentioned factors. The results confirmed a significant effect of following segment type ($F(1, 193) = 6.70$; $p = 0.010$) as well as an effect of L1 background ($F(2, 193) = 6.02$; $p = 0.003$) and speaker's gender ($F(1, 193) = 13.3$; $p < 0.001$).

In addition to analysing the effects of the type of neighbouring segments (consonant or vowel), we examined the influence of phonetic voicing in the neighbouring segments on the realisation of the function word *to*. The normalised word durations seemed to vary depending on the voicing of the following segment (115 ms vs. 138 ms in items with following voiceless vs. voiced segment). Figure 3.31 displays the normalised duration means for items followed by voiceless and voiced segments for the three speaker groups with different L1 backgrounds in the two speaking styles. An analysis of variance with the factors L1 background, speaking style and presence of phonetic voicing in the preceding and following segment confirmed a significant effect of following segment voicing on normalised duration ($F(1, 262) = 14.5$; $p < 0.001$). No other significant effects or interactions were found.

Furthermore, we inspected the effect of phonetic voicing in the neighbouring segments on segment proportions in the realisations of the word *to*. We found higher vowel proportions in items followed by voiced segments (23% vs. 28%, for items with voiceless vs. voiced right context). An analysis of variance with L1 background, speaking style and presence of phonetic voicing in the preceding and following segment as factors confirmed, apart from a significant effect of L1 background, the effect of following segment voicing ($F(1, 261) = 4.30$; $p = 0.039$). No other effects or interactions reached significance. As for the effect of phonetic voicing in the neighbouring segments on friction intensity in the plosive release phase, an analysis of variance with the factors L1 background, speaking style and presence of phonetic voicing in the preceding and

following segment showed no significant effects or interactions apart from the effect of L1 background.

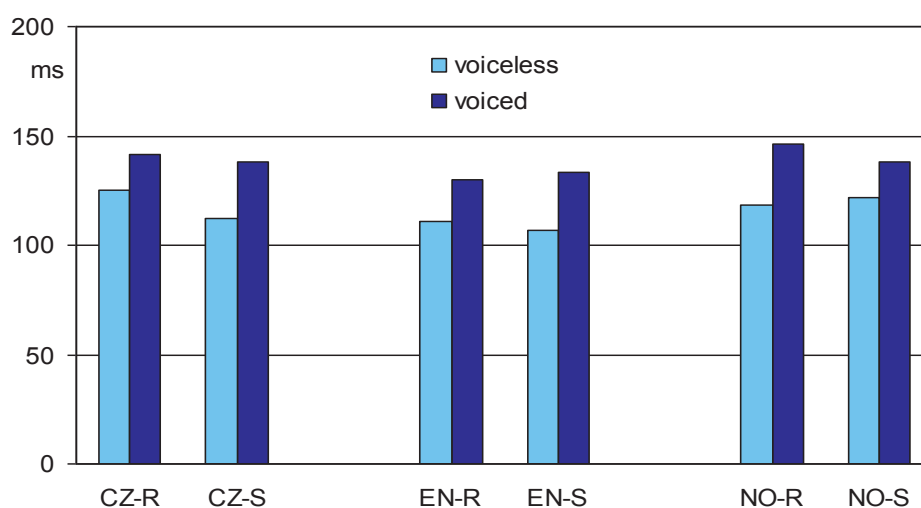


Figure 3.31: Normalised duration means (in ms) in the word *to* followed by voiceless/voiced segment, for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

As in the previous sections, here we will also describe the distribution of the various aspects of context in the present dataset. First we inspected the distributions of segment types of the immediately neighbouring left and right context, comparing the groups of items based on the speakers' L1 background and speaking style (see Tables 3.4 and 3.5 on page 98). Pearson's Chi-squared tests showed that neither the left segment type distributions nor the right segment type distributions varied significantly between the groups ($\chi^2(5) = 10.6$; $p = 0.06$ and $\chi^2(5) = 10.2$; $p = 0.071$, for left and right segment type distributions, respectively). Also the distribution of voiced/voiceless segments preceding the word *to* was not found to differ significantly for the groups based on the speakers' L1 background and speaking style ($\chi^2(5) = 3.02$; $p = 0.697$). The distributions of phonetically voiced/voiceless segments following the tokens of the function word *to*, however, differed significantly between the groups ($\chi^2(5) = 20.2$; $p = 0.001$). The distributions of phonetically voiced/voiceless segments following the word *to* for

groups of items based on the speakers' L1 background and speaking style are shown in Figure 3.32. As in the following segment voicing distributions for the word *of*, the number of tokens with voiceless following segments is higher among the English items compared to items produced by non-natives. This difference is very noticeable, especially in spontaneous speech. Due to that, we have to consider the possibility of confounding the effects of the L1 background and speaking style with the effects of unevenly distributed voicing in right segment (see Section 3.4 below).

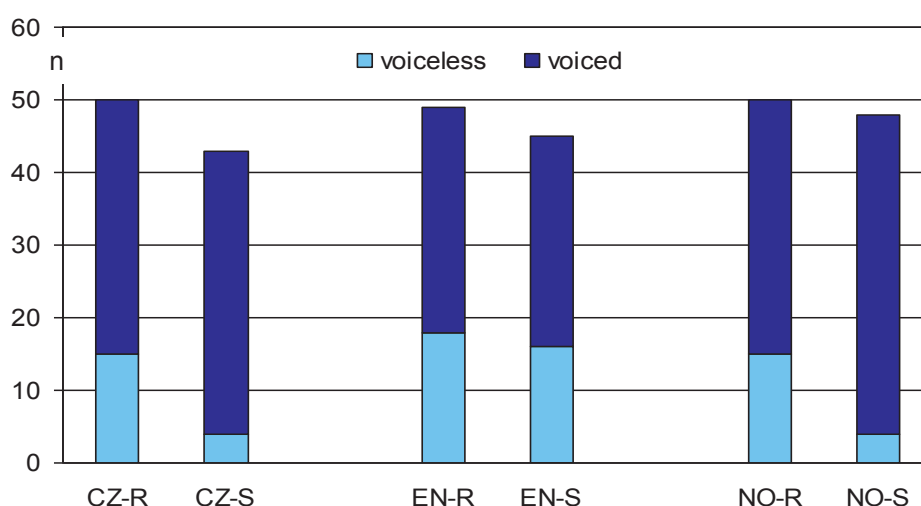


Figure 3.32: Distributions of phonetically voiced/voiceless segments following the word *to* (numbers of tokens), for speaker groups with different L1 backgrounds (CZ=Czech, EN=English, NO=Norwegian) and different speaking styles (R=read, S=spontaneous).

3.4 Summary and discussion

In the previous sections we have listed all the results found for each of the three function words separately. In this section, the results will be summarised and discussed with the relevant literature. Section 3.4.1 will summarise the main context effects on the function word realisations found in our material. As we have shown in the previous sections, the distributions of context voicing in segments neighbouring with the

observed words is not even for the items produced by speakers with different L1 backgrounds and in the two speaking styles. Therefore, we have to be aware of the significant context effects and consider the possibility that the observed main effects may be confounded with these context effects. Section 3.4.2 will then summarise main results relating to the temporal organisation of the function words, including the measurements of the articulation rate. The measures relating to the spectral properties of speech sounds will be summarised in Section 3.4.3.

3.4.1 Context influences

In this section we will summarise the effects of context on the realisations of function words. Section 3.4.1.1 will focus on the effects of neighbouring segment types on measures relating to temporal organisation, and on measures of vowel openness (only observed in the words *of* and *to*). In Section 3.4.1.2, the effects of the preceding sound's identity as nasal (for the word *in*), and phonetic voicing in the immediately neighbouring segments (for the words *of* and *to*) will be summarised.

3.4.1.1 Effects of neighbouring segment type

Table 3.9 presents an overview of the results of the statistical analyses relating to the effects of neighbouring segment types on measures of temporal organisation, and on vowel openness (in the words *of* and *to*). The cells corresponding to measures that were not observed for a given function word are crossed out, while the cells where the statistical tests did not find any significant effects on the observed measure are marked with “Ø”. In the following paragraphs, all the statistically significant results will be discussed.

When inspecting the first variable related to temporal organisation, the normalised word duration, some influence of the sound type of one of the neighbouring segments was found in all three observed function words. In the function words *of* and *to*, it was the following segment's type that had a significant effect on normalised durations, with words followed by consonants having shorter durations than those followed by vowels.

Chapter 3

The way the duration of the preposition *in* was affected by neighbouring segment types differed from this pattern. It was the type of preceding segment that had a significant effect, words preceded by consonants being shorter than those preceded by vowels.

Table 3.9: Overview of results of statistical analyses relating to the effects of neighbouring segment types on relevant measures of the realisation of the function words *in*, *of* and *to*.

	IN	OF	TO
Normalised duration	Left segment type *	Right segment type *	Right segment type *
Vowel proportion	Right segment type *	∅	Right segment type .
Fricative voicing proportion		∅	
Plosive release proportion			∅
F1 - F0 (with factor gender)		(Left segment type *)	Right segment type * (Right segment type *)

The results for the function words *of* and *to* are in agreement with previous research showing generally shorter durations, as well as a more frequent occurrence of other phenomena associated with phonetic reduction, in words followed by consonants, as compared to words followed by vowels (Jurafsky et al., 1998; Bell et al., 2003; van Bael et al., 2007). Jurafsky et al. (1998) investigated the effects of several factors on the durational and qualitative reduction of ten frequent English function words. Their results for the function words *of* and *to* show a considerable effect of the following segments' type (consonant or vowel) on word durations and the qualitative reduction of the vowels, as well as on the probability of coda deletion in the word *of*. For the word *of*, the duration was found to be 1.5 times longer before a vowel than before a consonant, and the word *to* was found to be 1.3 times longer when followed by a vowel

than when followed by a consonant. In contrast to that, the preposition *in* was found to be only 1.1 times longer before a vowel than before a consonant (which was, due to the amount of data used in their study, still significant). This is also in agreement with our results that showed an effect of following segment type on the duration of the words *of* and *to*, but no significant effect in the preposition *in*. The longer durations found for the word *in* following a vowel could be partly explained by a frequent occurrence of creaky voice quality on vowel-vowel transitions. Due to our segmentation criteria (cf. Section 3.2.3), when a period of creaky phonation occurred at the transition between a preceding word and the observed preposition, a part of this period was usually included in the vowel segment of the observed preposition. This may have contributed to the increased overall word duration in some cases.

Regarding the measure of vowel proportion within the words, we revealed quite different tendencies for the three words. While neither the left nor right context type had a significant effect on vowel proportion in the word *of*, vowel proportions in both *in* and *to* were found to be affected by the following segment's type (although the significance of this effect was only marginal in the word *to*). In both of these words, vowel proportion was lower when the words were followed by consonants than when they were followed by vowels. In the function word *to*, the items followed by a vowel had both longer plosives and longer vowels than items followed by a consonant (which obviously resulted in overall longer durations of items followed by vowels; see above). The difference in the vowel duration was, however, more noticeable. In the word *in*, vowel durations did not vary depending on the context, and nasals were considerably shorter before a vowel than before a consonant. The mentioned context effect in the function word *to* is, again, in agreement with Jurafsky et al. (1998), assuming that the more reduced vowel quality preceding a consonant also implies vowel shortening. The fact that there was no significant effect of the right context type on the vowel proportion in the function word *of* might be explained by the counter-acting effects of vowel shortening and more frequent coda deletion, both associated with following consonants (cf. Jurafsky et al., 1998), on the resulting vowel proportion.

Chapter 3

Further, as we can observe in the overview of results in Table 3.9, we revealed no significant effects of the neighbouring segment type on the other observed measures of segment proportions: the proportion of voicing in the fricative in the preposition *of*, and the proportion of release phase in the plosive in the word *to*.

In both words where formant measurements were made, we found the effects of segment type (of the segment directly neighbouring with the vowel in the observed word) on the degree of vowel openness. In the word *of*, the formant values were influenced by the preceding sound type, while in the word *to* it was the following segment type that had an effect on the degree of vowel openness. In both function words, the F1 - F0 Bark distance values were significantly higher in items where the observed vowel neighbored with a vowel, than when the neighbouring sound was a consonant. This consistent effect can be interpreted simply as a physiological influence of the different degrees of openness characteristic for the basic sound classes (vowel vs. consonant) on the first formant of the neighbouring vowel.

As was shown in the previous sections describing the context effects on the three function words, the distributions of segment types of the preceding and following segment were not found to differ significantly between the groups of items produced by speakers with different L1 backgrounds in different speaking styles in any of the three words. Therefore, the above described significant context effects are not likely to distort the results of the analyses of L1 and speaking style effects.

3.4.1.2 Other context effects

Table 3.10 presents an overview of the results of the statistical analyses relating to the effects of the preceding segments' identity as nasal (for the word *in*) and the neighbouring segments' phonetic voicing (for the words *of* and *to*), on the relevant temporal and spectral measures. Apart from the context influences on normalised durations and vowel proportions, we inspected the effect of the preceding segments' identity as nasal on the degree of nasalisation of the vowel in the word *in*, and on the

degree of nasalisation of the segment immediately following the preposition *in*. Further, we inspected the effects of the neighbouring segments' phonetic voicing on fricative voicing proportion and friction intensity in the fricative (in the word *of*), and on the friction intensity in the plosive release (in the word *to*). The cells corresponding to measures that were not observed for a given function word are crossed out, while the cells where the statistical tests did not find any significant effects on the observed measure are marked with "Ø". In the following paragraphs, all the statistically significant results will be discussed.

Table 3.10: Overview of results of statistical analyses relating to the effects of neighbouring segments' identity as nasal (for the word *in*) and neighbouring segments' phonetic voicing (for the words *of* and *to*) on relevant measures of the realisation of the function words.

	IN	OF	TO
Normalised duration	L1:left nasal *	Left voicing **	Right voicing ***
Vowel proportion	Left nasal **	Right voicing *	Right voicing *
Fricative voicing proportion		Right voicing ***	
Vowel nasalisation degree	Left nasal ***		
Following segment's nasalisation degree	Ø		
Friction intensity in fricative		Right voicing ***	
Friction intensity in plosive release			Ø

Chapter 3

The first column in Table 3.10 lists the effects of the preceding segments' identity as nasal on the relevant variables observed in the realisations of the preposition *in*: normalised duration, vowel proportion, vowel nasalisation degree and the degree of nasalisation in the segment following the preposition. As for the normalised word duration, we observed a significant interaction of the context and the speakers' L1 background. There was a strong tendency for longer word durations following a nasal in items produced by native English speakers, and still considerably longer durations after a nasal in items produced by Norwegian speakers. Czech speakers, however, showed a weak opposite tendency. Further, vowel proportion was found to be affected by the preceding segment's identity as nasal. The items following a nasal had higher vowel proportions than words following a non-nasal segment. This was mainly due to the increased duration of vowels, while the duration of nasal was only slightly shorter in the words following a nasal. In addition, a large effect of the preceding segment's identity as nasal was found on the auditorily determined degree of vowel nasalisation. The fact that vowels preceded by nasals (in fact *surrounded* by nasals, since all observed vowels were also followed by nasals in the word *in*) were produced with a significantly higher degree of coarticulatory nasalisation is not surprising. Perhaps more interestingly, there was no interaction with the L1 background. This indicates a comparable amount of nasal coarticulation for the three speaker groups. The data therefore support the assumption that since the L1 of all involved speaker groups has no phonological distinction of nasality in vowels, the coarticulatory nasalisation of vowels takes place to a similar degree in all speaker groups. In contrast to the large effect of the preceding segment's phonological identity as nasal on the vowel in the preposition *in*, we did not find any effect of this factor on the degree of nasalisation of the segment following the word *in*.

As to the effects of phonetic voicing in the neighbouring segments that we observed in the function words *of* and *to*, the normalised durations varied considerably depending on the phonetic voicing of the neighbouring segments. For the word *of*, it was the preceding segment's voicing, while the word *to* was influenced by the phonetic voicing in the following segment. In both words, the normalised durations were shorter in the

neighbourhood of a voiceless segment. Vowel proportion in the function words also seemed to be influenced by the presence of phonetic voicing in neighbouring segments. In both *of* and *to*, the vowel proportion was higher in words followed by voiced segments. In the word *to*, we could observe that vowel durations are considerably longer when followed by a phonetically voiced segment. The duration of plosives in such cases was also longer, although this was less noticeable. In the preposition *of*, we could observe slightly longer fricative durations before a voiceless following segment, while the vowel durations were shorter in these contexts. The pattern found in the function word *to* is consistent with a well-known effect of voicing in the following consonant on vowel duration (e.g. House and Fairbanks, 1953; House, 1961; Kluender et al., 1988; van Santen, 1992) although it should be noted that our material is not controlled for stress, utterance position and other factors that were found to interact with this effect. The pattern found in the word *of* seems rather complex and less clear to interpret. In addition to the previously discussed temporal measures, we expected a large influence of neighbouring segments' phonetic voicing on the two measures related to voicing (fortis/lenis character) of the fricative in the preposition *of*: the fricative voicing proportion and the friction intensity in the fricative. As expected, the analyses revealed a large influence of the phonetic voicing in the segment immediately following the fricative on both measures. The voicing proportion was greater and the friction intensity lower in fricatives followed by voiced segments, consistently in the three speaker groups with different L1 backgrounds. Moreover, this effect appeared to be stronger in spontaneous than in read speech. It can be speculated that this consistent effect of the following segment voicing is due to similar regressive voicing assimilation processes existing in the three languages (cf. Section 1.3.3). As opposed to the effects of neighbouring segments' phonetic voicing on the observed measures related to voicing in the word *of*, the measure of friction intensity in the release portion of the plosive in the word *to* was not found to be affected by the presence of phonetic voicing in either the preceding or following segment.

Apart from describing the effects of some of the context properties, we also inspected the distributions of context with regard to the segments' identity as nasal (for the word

Chapter 3

in) and the neighbouring segments' phonetic voicing (for the words *of* and *to*), in the groups of items based on the speakers' L1 background and speaking style. The distribution of nasals preceding the preposition *in* was found not to differ between the groups. Similarly, the distribution of the voiced and voiceless preceding segment did not vary significantly between the groups based on the speakers' L1 background and speaking style in either *of* or *to*. In contrast to that, the distributions of phonetic voicing in the following segment in both the function words *of* and *to* differed significantly. In both function words, we found that native English speakers had significantly larger proportions of voiceless following contexts. In the word *to*, the difference between speaker groups was more noticeable in spontaneous speech, where both non-native groups had extremely low proportions of voiceless following contexts. These findings are very important to keep in mind when interpreting the results of measures that were shown to be strongly affected by the following segment voicing. Based on the previous analyses, this would include the normalised durations in the word *to*, the vowel proportions in both *of* and *to*, and especially the voicing proportion and friction intensity in the preposition *of*, which will be discussed in the next sections.

3.4.2 Temporal organisation

Table 3.11 presents an overview of the results of the statistical analyses relating to variables describing the temporal organisation of the three investigated function words. The cells corresponding to measures that were not observed for a given function word are crossed out, while the cells where the statistical tests did not find any significant effects on the observed measure are marked with “Ø”. In the following paragraphs, all the statistically significant results will be discussed.

The first variable that relates to the overall temporal properties of the speakers' production is the articulation rate, which was measured in the speech stretches surrounding the observed function words (in syllables per second). In our sample, we found significantly higher articulation rates in items spoken by the native speakers as compared to both the non-native groups, as well as a difference between the articulation rates in the items produced by the Czech and the Norwegian speakers. The Norwegian

speakers' articulation rate was still significantly higher than the rate of the Czech speakers. In addition, there was an effect of speaking style. The articulation rates were on average higher in read speech, in all three speaker groups. The absence of an interaction of speaking style and L1 background indicates that the differences between the production mechanisms associated with the two speaking styles manifest similarly in all three speaker groups with different L1 backgrounds.

Table 3.11: Overview of results of statistical analyses relating to variables describing the temporal organisation of the function words *in*, *of* and *to*.

	IN	OF	TO
Articulation rate	L1 *** Style ***		
Normalised duration	∅	L1 *** Style * L1:style	L1 **
Vowel proportion	∅	L1 ** Style *** L1:style **	L1 ***
Fricative voicing proportion		L1 *** Style *** L1:style	
Plosive release proportion			L1 *** Style ***

The consistently slower articulation rates in spontaneous speech may also be partly explained by the higher cognitive demands associated with the replication task used for the elicitation of spontaneous speech. Moreover, the articulation rate variability was higher in spontaneous speech. The slower articulation rates for both non-native groups are in agreement with a number of studies that investigated speech rate or matched sentence durations as measures of fluency in non-native productions. Overall, productions by non-native speakers or speakers with less experience in a given language (e.g. less experienced vs. more experienced L2 speakers, late vs. early bilinguals) were

Chapter 3

shown to have lower speech rate or articulation rate, or longer durations of matched sentences than native, or more experienced speakers of the language (e.g. Riggenbach, 1991; Towell et al., 1996; Guion et al., 2000; Cucchiariini et al., 2002; MacKay and Flege, 2004; Kormos and Dénes, 2004; Trofimovich and Baker, 2006; Toivola et al., 2010; for details see the literature overview in Section 1.2.2). As mentioned in Section 2.2.1, there were some differences in the variables relating to the non-native speakers' experience with English between the Czech and Norwegian speaker group. Moreover, we could speculate about the overall lower exposure to spoken English in the Czech Republic as compared to Norway. It is possible that the slower articulation rates of the Czech speakers compared to those of the Norwegians in this data set are due to these differences related to L2 experience, which would be consistent with previous research (cf. Towell et al., 1996; Guion et al., 2000; Cucchiariini et al., 2002; MacKay and Flege, 2004; Trofimovich and Baker, 2006). On the other hand, an evaluation of the non-native speakers' fluency using considerably longer random samples of spontaneous speech showed particularly a large between-speaker variation within each of the speaker groups (for details see Section 2.2.2).

As to our results relating to the articulation rate differences between the two speaking styles, the higher articulation rate in read speech in our data is in accordance with previous studies using similar types of speech material (Hirschberg, 2000; Mixdorff and Pfitzinger, 2005; for details see the literature overview in Section 1.2.1.2). On the other hand a number of studies have found an opposite tendency, i.e. higher articulation rates in spontaneous speech (cf. Section 1.2.1.2). We believe that the different tendencies in various studies can be explained by a low number of speakers in some of the studies, as well as by differences in the conditions at which the spontaneous speech was produced. Moreover, our results showing a greater variation coefficient of articulation rate in spontaneous speech are consistent with a study of Koopmans-Van Beinum (1992) which showed a larger speech rate range and variability in spontaneous speech.

The next inspected temporal measure, normalised word duration, reports on the durational relation of the observed words to the surrounding speech rather than on the

actually measured duration, which is significantly correlated with the articulation rate. The normalisation, described in detail in Section 3.3.1.1, removes the variance due to the articulation rate, and the obtained values represent a hypothetical duration, that *would be produced at the mean articulation rate*. In tokens produced at high articulation rates, the normalised word durations were longer compared to actually measured durations (“stretched”), while in items produced in slow speech the normalised word durations were shortened compared to the raw durations (“compressed”). The results showed that whereas in the word *in*, no differences in normalised durations were found between the speaker groups with different L1 backgrounds, for the words *of* and *to*, the normalised durations still differed between the speaker groups. More precisely, in the word *of*, the normalised durations were significantly longer in the items produced by the Czech speakers, as compared to both the native speakers’ normalised durations and those produced by the Norwegians. In the word *to*, on the other hand, the natives had significantly shorter normalised durations than both non-native groups. Regarding the speaking style, there was no effect on the normalised durations of the words *in* and *to*, but for the word *of*, the mean normalised duration was shorter in spontaneous speech. Moreover, there was a significant interaction of the L1 background and speaking style effects: while the Czech speakers’ read items were considerably longer than their spontaneous ones, the difference was smaller for the Norwegian speakers, and native speakers even showed a slightly opposite pattern.

As a first possible reason for the natives’ significantly shorter normalised durations of the word *to*, we should recall that a significant effect of phonetic voicing in the following segment on normalised durations was found, and the distribution of voiced and voiceless segments following the word *to* was found to differ between the speaker groups with different L1 backgrounds. In particular, the native speaker group had higher proportions of voiceless following contexts than both non-native groups. This was especially noticeable in spontaneous speech, due to the very low numbers of items with voiceless following segments in the two non-native groups (see Section 3.3.4.3 for more details). Moreover, an analysis of variance with the factors L1 background and speaking style, as well as preceding and following segment’s voicing, no longer found a

Chapter 3

significant effect of L1 background. Although the uneven distribution of right context voicing in the three speaker groups may partly explain the normalised duration differences, the mean values still vary noticeably after controlling for context voicing, albeit the difference is no longer significant with the smaller group sizes. We can therefore still assume that the L1 background effect is not merely an artefact of the uneven distribution of voicing in the following segment. In sum, the L1 effect on normalised word durations of the words *of* and *to* indicates that in these two words there may exist additional reduction mechanisms beyond the influence of the articulation rate that determine the word durations, and these mechanisms are not fully mastered by non-native speakers. In the word *in*, on the other hand, any durational differences between the native and non-native productions may be attributed purely to the articulation rate. A possible explanation for this pattern of results would be the existence of weak forms of certain English function words (e.g. Jones et al., 2003: 589; Roach, 1983: 86-93). While in the pronunciation of the words *of* and *to*, weak forms are described in addition to the full (strong) forms, the preposition *in* is usually not included among such weak-form words (Jones et al., 2003: 271, 377, 539; Roach, 1983: 86-93). Although the tokens in the investigated sample did not include cases where a strong word form would be expected (as for example in clause-final position; see Section 3.2.2 for more details about the selected tokens), it may be assumed that non-native speakers have a lower awareness of weak form use and tend to use the strong forms in a wide range of contexts instead. This explanation would be consistent with the observed longer normalised durations of the weak form words *of* and *to* in non-native as compared to native production.

The next measure relating to temporal organisation that inspected was the vowel proportion within the observed words. It needs to be remembered that due to the different structures of the three observed words, this measure cannot be expected to show consistent tendencies across the three words. While we found no effect of the speakers' L1 background on vowel proportion in the word *in*, in both *of* and *to*, there were significant differences in vowel proportion between the speaker groups. In productions of the word *of* it was the group of Norwegian speakers that differed from

the rest of the speakers, having higher vowel proportion. The vowel proportions of the natives and the group of Czech speakers did not differ significantly. In the word *to*, on the other hand, the productions of native speakers differed from both non-native groups by lower vowel proportions. In addition to the effect of L1, vowel proportions in the word *of* varied also depending on speaking style, the mean values being lower in read speech (45%) than in spontaneous speech (52%). This effect was, moreover, in interaction with the L1 background. While the natives showed a moderate effect, this effect was much larger for the Czech speakers. On the contrary, the group of Norwegian speakers showed a weak opposite tendency to the mentioned trend.

First, we should note that a significant effect of phonetic voicing in the following segment on vowel proportions in both *of* and *to*, was found in previous analyses. Since we also confirmed that native speakers had a higher number of items with voiceless following segments than the non-native speaker groups in both *of* and *to*, we need to consider the possibility that the vowel proportion differences between the speaker groups might be partly due to this uneven context distribution. A closer inspection of the results shows, however, that the effect size of the L1 background is larger than the possibly confounding effect of right context voicing, and other explanations for the L1 differences in vowel proportions should be sought. There may be a simple explanation for the dissimilar patterns of L1 influence in the words *of* and *to*. The native-like vowel proportions of Czech speakers in the productions of the word *of* may result from their very long durations of fricatives (particularly in read speech). When we inspect raw durations of vowels in the three speaker groups, we find that it is in fact the native speaker group that has noticeably shorter vowel durations than both the Norwegian and Czech speakers. But since the Czech speaker group also produced unusually long fricatives, their overall vowel proportion does resemble the ratio of the native speakers. The function word *to* does not offer such variability in the duration of the consonantal segment, and the values of vowel proportion thus reflect more directly the actual temporal relations. The main difference between the natives and both non-native groups was, as in the case of the word *of*, the natives' shorter vowel duration. On the contrary, no difference in vowel proportion between the native and non-native productions was

Chapter 3

found for the word *in*. In sum, native productions of the words *of* and *to* tend to have shorter vowels, which is reflected by lower vowel proportions, as compared to those of non-native productions (with the exception of Czech productions of the word *of* in read speech, characterised by unusually long fricatives). We may speculate that this is the result of the non-natives' inability to reduce vowels below a certain duration. The native speakers' drastic reduction of the vocalic element in the word *to* may also be seen as a result of processes described as schwa absorption (Shockey, 2003: 22-26). In the case of voiceless stops, Shockey argues that the syllabic property of the schwa overlaps with the articulatory quality of the stop. Apparently, non-native speakers do not apply such processes to a comparable degree. In the case of the preposition *in*, on the other hand, we may assume that the structure of the word, or its lack of a weak form, makes it less prone to be affected by language-specific reduction processes. As for the speaking style effect on vowel proportion and the interaction of the L1 background with speaking style (found in the preposition *of*), we should again mention the unusually long fricatives in the Czech speaker's read items. This may be one of the reasons for the seemingly strong effect of speaking style, as well as the strong interaction. The higher vowel proportion in spontaneous speech may also be explained by the [v/0] alteration, that Shockey (2003: 34-35) describes as a connected speech process contributing to preserving a CV-type syllable structure. The weak form [ə] of the word *of* is typically realised in casual speech when the word is followed by a consonant.

The proportion of voicing in the fricative in the preposition *of* showed, just as the previous temporal measures in this word, to be affected both by L1 background and speaking style. A significant interaction of these two factors was found as well. Similarly as with the vowel proportion, also in the proportion of fricative voicing the difference was found between the productions of Norwegian speakers and the remaining two groups. The Norwegians produced the fricatives significantly more voiced (85%) than both the native speakers (65%) and the Czech speakers (72%). The overall mean voicing proportion in read items was 65% while in spontaneous speech it was 83%. However, the differences, varied across the speaker groups with different L1 backgrounds. While both the natives and the Norwegian speakers had a smaller

difference between the voicing proportion in read and spontaneous speech (62% vs. 68% for natives, and 80% vs. 90% for Norwegians), the speaking style effect in the Czech speaker group was very large: 56% vs. 90%.

Here too, it needs to be repeated that the very long durations of the Czech fricatives in read speech may be responsible for some of the differences. While the voicing proportion in the spontaneous Czech items was as high as the Norwegians', their read items have a mean voicing proportion lower than that of the natives. Since the mean duration of the phonetically voiced portion is comparable with the Norwegians' values, the very low voicing proportion in the Czech read speech seems like a mere consequence of longer fricative durations. On the other hand, the lower values of native speakers were not due to unusually long fricatives, but simply reflected the noticeably shorter mean duration of voiced portions of the fricatives. In addition, there is another circumstance that we should take into consideration when explaining the differences in fricative voicing proportion among the groups based on the speakers' L1 backgrounds. The distribution of voiced and voiceless following segments was uneven in the speaker groups with different L1 backgrounds, and since this context factor proved to have a strong influence on the amount of fricative voicing, the difference between the speaker groups may have been a result of this uneven context distribution (in particular a higher number of tokens with voiceless following context among the native speakers' tokens; for more details see Section 3.3.3.3). This explanation was further supported by an analysis of variance with the factors L1 background and speaking style, as well as the preceding and following segment's voicing. Here, the effect of the factor L1 background was no longer significant. It is, however, difficult to state with certainty, if the difference in fricative voicing proportion between the speaker groups with different L1 backgrounds was a mere artefact of the uneven distribution of voicing in the following segment. We have to be aware that the analysis of variance with four factors is weaker because the data are split into a much higher number of cells with smaller number of observations.

Chapter 3

Apart from these incidental reasons, the differences in fricative voicing proportion between the speaker groups can be explained by the different phonological systems of English and of the native languages of the two non-native groups, in particular the phonetic properties associated with phonological voicing contrast and the phonological processes related to voicing. Shockey (2003: 30-31) mentions that in English, a certain amount of phonetic devoicing is expected, since the phonological voicing contrast is signalled by other means such as preceding vowel length. The phonetic correlates of phonological voicing in Norwegian include, as in English, the duration of the segment and the duration of the preceding vowel, as well as the presence of aspiration in some positions (cf. Section 1.3.3). But since segment duration in Norwegian also functions as a cue to the phonological length of vowels, we may assume that its use to signal voicing is somewhat more limited. At the same time, no final devoicing is present in Norwegian, and therefore, the (phonological) voicing distinction is maintained in word-final positions (e.g. Kristoffersen, 2000: 74-75; Husby and Kløve, 2001: 62). It should be pointed out, however, that no research focussed particularly on the presence of phonetic voicing in word-final consonants. In contrast with English and Norwegian, voicing distinction in Czech is mainly signalled by vocal fold vibration, although differences in the duration of voiceless and voiced sounds (in particular, longer durations of voiceless sounds) have also been observed (e.g. Palková, 1997; Machač and Skarnitzl, 2007). In addition, the phonological voicing contrast in Czech is neutralised in pre-pausal positions as well as in obstruent clusters. More details about voicing contrast and voicing assimilation in the studied languages can be found in Section 1.3.3. The above mentioned facts might indicate that whereas native English speakers may produce phonetically partly devoiced fricatives while signalling the phonological voicing with other means, and Czech speakers may fail to produce fully (phonetically) voiced sounds as a result of the transfer of rules applied in their native language, the Norwegians' productions show much smaller amounts of devoicing, which is also in agreement with the relevant aspects of the phonological system of their native language. However, a study containing more material would be needed to better explain the results relating to this issue.

It is possible that the differences in voicing proportion in the two observed speaking styles are also partly due to the previously mentioned very long fricatives in the Czech read speech. As has been explained above, the long durations of fricatives may be the main reason for the very low voicing proportions in these items, which contributes to the overall speaking style effect, as well as to the significant interaction of the L1 background and speaking style. As confirmed by raw durational values, the higher voicing proportion in spontaneous speech reflects both the shorter durations of fricatives and longer duration of voicing phase in spontaneous speech, present consistently in all the speaker groups with different L1 backgrounds. Both the durational reduction and the tendency to fricative lenition are generally associated with less careful speech (Lass, 1984: 177-183; Barry and Andreeva, 2001; Shockey, 2003: 27-28). Furthermore, this result is consistent with the findings by Van Son and Pols (1999). This study compared various parameters in consonants produced in read and spontaneous speech in Dutch and revealed, among other things, higher values of the centre of gravity in fricatives in read as compared to spontaneous speech, which correspond to a more fortis character of fricatives. This is in turn associated with greater articulatory effort during fricative production in read speech.

The last inspected measure relating to the temporal organisation was the proportion of release phase in the plosive in the function word *to*. As in vowel proportion measure in the same word, in the plosive release proportion we also found an effect of the L1 background, with native speakers' values differing significantly from both non-native speaker groups. The proportion of release phase duration within the duration of the plosive was higher in the items produced by natives. In addition, there was an effect of speaking style: the items produced in read speech had consistently smaller proportions of release phase. It should be mentioned that in most cases the quality of the release indicated mainly the presence of local friction rather than aspiration. This is in agreement with the expectation of less noticeable aspiration in unstressed syllables (e.g. Davidsen-Nielsen, 1977: 52).

Chapter 3

A higher release proportion in fact indicates a relatively shorter closure phase resulting from a reduction of articulatory effort in more casual speaking styles (Shockey, 2003: 27-28). This is in agreement with our results, which show consistently higher proportions of the release phase in spontaneous than in read speech across the speaker groups. The related observation of the more frequent occurrence of items without a complete plosive closure in spontaneous speech is also consistent with this explanation.

The higher release proportion (and lower closure proportion) in the native group compared to both non-native groups may also be the result of more frequent affrication of plosives in native speakers' productions. The affrication of voiceless plosives was described in several British English dialects, including colloquial speech in SSBE (Gimson, 1980: 160; Sangster, 2001; Jones and Llamas, 2003). Although non-native speakers may be aware of this phenomenon, it does not belong to the pronunciation standard that L2 instruction usually aims for, and therefore, it cannot be expected in non-native productions. In addition, in the phonological system of Czech, an alveolar affricate phoneme contrasts with an alveolar plosive, which further prevents the inadvertent confusion of the two sounds by Czech speakers. The fact that there was no interaction between the L1 background and speaking style indicates that the tendency to produce plosives with a slightly shorter closure phase in spontaneous speech is rather universal and does not cause additional problems for non-native speakers.

3.4.3 Spectral properties of speech sounds

Table 3.12 gives an overview of the results of the statistical analyses of measures relating to the spectral properties of speech sounds in the three investigated function words. The cells corresponding to measures that were not observed for a given function word are crossed out, while the cells where the statistical tests did not find any significant effects on the observed measure are marked with "Ø". In the following paragraphs, the statistically significant results will be discussed.

In the word *in*, the only measure relating to spectral properties was the auditorily determined degree of nasalisation of the vowel. In our results, we found neither an

effect of L1 background nor an effect of speaking style on this measure of vowel quality.

Table 3.12: Overview of results of statistical analyses relating to variables describing the spectral properties of speech sounds in the function words *in*, *of* and *to*.

	IN	OF		TO		
		female	male	female	male	
Vowel nasalisation degree	∅	/		/		
Following segment's nasalisation degree	L1 *** L1:style					
F1 - F0	/		L1 . Style *		∅	
F1			∅	L1:style *	L1 ***	L1 ***
F3 - F2	/		L1 ***		Style **	
F2			L1 ***	L1 ***	L1 ***	L1 ** Style *
Friction intensity in fricative	/		L1 *** Style *** L1:style		/	
Friction intensity in plosive release			L1 *** Style *			

In contrast to the absence of the effects of L1 background and speaking style on the degree of nasalisation in the word's vowel, we found a significant effect of the L1 background on the degree of nasalisation in the segment immediately following the preposition *in*, as well as an interaction of the L1 background with speaking style. The

Chapter 3

following sound's nasalisation measure does not refer directly to the realisation of the function word *in*, but it provides additional information about the integration of this word in context, evidenced by qualitative changes relating to the occurrence of a nasal. Therefore, it seems reasonable to discuss these results together with the qualitative aspects of the word's realisation. The results showed that the degree of nasalisation in the following segment was significantly lower in the items produced by the Czech speakers, compared to both the native speakers and the Norwegians. The interaction of the L1 background with speaking style was due to the Norwegians' differing behaviour in the two speaking styles. While the Czech and native English speakers had almost the same mean values of degree of nasalisation in a following segment for both read and spontaneous speech, for the Norwegians the mean values for the two speaking styles differed considerably. Their mean nasalisation degree in read speech was not much higher than that of the Czech speakers, but their mean value in spontaneous speech was even higher than the values produced by the natives.

The absence of effects of L1 background or speaking style on vowel nasalisation in the word *in* is not surprising if we consider that the native languages of all three speaker groups have a similar vocalic system in the sense that they contain only oral vowels. Unlike languages such as French or Portuguese, the languages in question do not distinguish between oral and nasal vowels. This may predispose them to a higher degree of nasal coarticulation of vowels than would be acceptable in languages where such a contrast is phonologically distinctive. The degree of nasalisation in the segment immediately following the preposition *in*, on the contrary, showed different patterns for the three speaker groups. The lower degree of nasalisation in the following segment in productions of the Czech speakers, as compared to the other two speaker groups, may be a result of an L1-specific tendency to indicate word boundaries by reducing coarticulation across word boundaries. A different phenomenon illustrating this tendency in the English productions of Czech speakers was described by Bissiri and Volín (2010). They found that the glottalisation of word-initial vowels in Czech English occurs frequently in all positions in relation to prosodic structure. Native English speakers, on the other hand, seldom glottalise at non-phrase boundaries, and even at

phrase boundaries they glottalise far less frequently than Czech speakers. We may speculate that Czech speakers' reduced nasal coarticulation across word boundaries is a related phenomenon aimed at increasing intelligibility by indicating word boundaries with various means.

In both words where we performed formant measurements (*of* and *to*), significant differences were found between the mean values for the speaker groups with different L1 backgrounds. In the word *of*, we found a highly significant effect of L1 background on Bark distance F3 - F2 (which was confirmed by analyses of raw F2 values separately for females and males). Here, the group of native speakers had significantly higher F2 values (lower F3 - F2 values), corresponding to more fronted vowel articulation. In the word *to*, on the other hand, we initially did not find any effect of L1 background on either of the observed Bark distance measures. However, further analyses inspecting raw formant values separately for males and females did reveal L1 background effects. It turned out that in both male and female groups, the native speakers had significantly higher F1 values, indicating more open vowel articulation. Moreover, we could observe higher variability of F1 - F0 values (or F1 values, observed separately for females and males) in native productions compared to the productions of both non-native groups. In addition, the analyses of raw F2 values revealed significant effects of L1 background, both in female and male subsets of data. The results differed somewhat between females and males. For females, the values of the Czech speakers were similar to the Norwegians' values (i.e. both significantly lower than the natives' values). For males, on the other hand, the Norwegians had the highest mean F2 value, but only differed significantly from the Czech values. We can summarise, that the vowel quality in both function words differed significantly between the native group and the non-natives' productions. In the word *of* the difference was due to more fronted pronunciation of the natives, while in the word *to*, the main difference was the higher degree of vowel openness in natives. The other differences were not consistent between females and males, and may be due to rather small number of speakers (when separating the two genders). Interestingly, there were almost no effects of speaking style, particularly when taking into account speaker gender. In the word *of*, we found a speaking style effect on

Chapter 3

the F1 - F0 measure, indicating slightly more open vowel qualities in read speech. Further analyses of raw formants in the male and female speaker groups separately, however, did not confirm this tendency. In the word *to*, on the other hand, the F3 - F2 measure showed significantly higher values in spontaneous speech than in read speech, indicating more fronted vowel qualities in read speech. When testing females and males separately, we observed that the males had higher F3 - F2 values in spontaneous speech (corresponding to significantly lower F2 in spontaneous speech) and although data from females showed a similar pattern, the speaking style effect in the female data was not significant.

The fact that vowel quality in the function words *of* and *to* differs between native and non-native productions is in agreement with general expectations about L2 productions. A number of studies described inadequate vowel qualities in non-native production and noted the effect of L2 experience on vowel production. For instance, Bohn and Flege (1992) examined the accuracy of vowel production by early and late German-English bilinguals. Formant measurements showed that their more experienced speakers produced English /æ/ more accurately than the less experienced late bilinguals. Flege et al. (1997) observed inappropriate spectral contrasts in English vowel pairs produced by German, Mandarin, Spanish and Korean speakers. The vowels produced by less experienced speakers were less native-like than those of more experienced speakers. However, in contrast to our investigation, the above mentioned studies focussed only on the production of vowels in stressed syllables (in monosyllabic content words). A very relevant study of the production of unstressed English vowels by early and late Korean-English and Japanese-English bilinguals was carried out by Lee et al. (2006). In this study, formant measurements showed that both groups of late bilinguals produced vowels more dispersed in the vowel space, with formant frequencies similar to full vowels with the same orthographic representation. The group of early Japanese-English bilinguals showed a weaker similar tendency. In contrast to that, the early Korean-English bilinguals had the smallest dispersion for English unstressed vowel production which suggested that they use a native Korean vowel target /ɨ/ (while the Japanese vowel system does not have any high- or mid-central vowels). Consistently with the

findings of Lee et al. (2006), our spectral results may be explained by non-natives' tendency to use full vowels typical for strong forms of the particular function words. The present results showed less fronted vowel quality in the non-native productions of the word *of*, (characteristic for its strong form vowel /ɒ/) as well as more closed vowel quality in the non-native productions of the word *to* (characteristic for its strong form vowel /u:/). The native speakers, on the other hand, produced more centralised, schwa-like vowels in both function words. The spectral results then apparently support the speculation that non-native speakers have lower awareness of weak forms of certain English function words (cf. results relating to temporal organisation discussed in Section 3.4.2). Moreover, the higher F1 variability in the function word *to* observed in the native speaker group (particularly among the females) may indicate a higher degree of coarticulation of the vowel, compared to the non-natives' production aiming more strenuously for a specific target quality. This is in accordance with research showing that the quality of schwa is strongly influenced by its segmental context (e.g. Browman and Goldstein, 1992; Kondo, 1994). Kondo (1994) investigated schwa variation in English, using systematically varied contexts. She found that the extensive variation of F2 values can be largely explained by coarticulation with its context, while the context effects on F1 are smaller. It has to be pointed out, however, that the speculation about the causes of native speakers' higher F1 variability cannot be verified based on the present data, since no detailed classification of articulatory settings in neighbouring segments is available for the investigated items, and the data sample is probably not of sufficient size for this kind of analysis. Regarding the Norwegians' higher F2 values in the word *to*, observed in items produced by males, a possible explanation is the Norwegians' tendency to substitute the English sounds /ʊ/ and /u/ with the Norwegian vowel /ɥ/ (e.g. Davidsen-Nielsen 1977: 92-93, 98).

The observed higher F3 - F2 Bark distance in spontaneous items as compared to read speech in the word *to* seemed to be the only relatively reliable speaking style effect on the inspected variables relating to the spectral properties of speech sounds. Higher F3 - F2 values indicate less fronted vowel articulation in spontaneous items, or vice versa, more fronted vowels in read speech, which does not seem to have any apparent

Chapter 3

explanation consistent with the reduction tendencies in more casual speaking styles. As mentioned above, this tendency (and the corresponding pattern of raw F2 values) was significant only for males. This, in sum, may indicate that this speaking style effect, although found to be significant in the present data, is less general.

In addition to the particular results describing vowel formant productions, our analyses brought a more general finding relating to methods for formant analysis. We originally intended to analyse vowel qualities using the Bark distance measures proposed by Syrdal and Gopal (1986). According to data presented by Syrdal (1985) and Adank et al. (2004), this method successfully reduces anatomical variation corresponding to gender. Table III in Adank et al. (2004), for example, indicates that Bark distance normalisation reduces gender specific variation to chance-level (i.e. linear discriminant analysis obtained results at chance-level). In addition, the Bark distance measures are widely used in research (e.g. Bohn and Flege, 1992; Flege et al., 1997; Baker and Trofimovich, 2005), since they are advantageous for analysing samples that contain data from both genders. In our data, however, we revealed that the Bark distance measures (especially the F1 - F0 measure) retain a significant between-gender variation. This urged us to perform additional analyses of single formant values in Bark (separately for males and females). In general, this finding suggests that the degree to which Bark distance normalisation reduces anatomical variation related to speaker gender may vary with the types of material used. Therefore, attention should be paid to checking normalisation characteristics on particular material.

As for the measures of friction intensity in the fricative in the preposition *of* and in the release phase of the plosive in the function word *to*, we found largely similar tendencies for both words. The language background proved to be highly significant, with Norwegians' productions having significantly lower friction intensity than both other groups. In the fricative in the word *of*, we also found that native speakers had still significantly higher friction intensity than Czech speakers. Also the effect of speaking style was found in both words. The tendencies, however, differed between the two function words. While friction intensity in the fricative in the preposition *of* was

considerably lower in spontaneous speech, in the word *to*, the plosive release friction intensity was lower in read speech. Moreover, there was an interaction of L1 background and speaking style in the word *of*.

First, we should repeat that the friction intensity in the fricative in the word *of* has a close relationship to the temporal measure of voicing proportion in the fricative, which was discussed in the previous section. It was confirmed that the friction intensity measure shows largely the same tendencies as the voicing proportion measure. Apart from the effects of the main factors L1 background and speaking style, friction intensity was also influenced by the presence or absence of phonetic voicing in the following segment. This fact, in combination with the uneven distribution of phonetic voicing in the following segment in the three speaker groups, obviously casts doubt on the validity of the determined effect of L1 background. However, unlike the temporal measure of voicing proportion, the measure of friction intensity retained a significant effect of L1 background even in a 4-way ANOVA containing the preceding and following segment's voicing as factors along with the main factors of L1 background and speaking style. The speaking style effect, however, did not reach significance in this analysis. As we discussed in Section 3.4.2, the three languages in question differ in phonetic characteristics signalling the phonological voicing contrast, as well as in some phonological processes related to voicing, which may explain different amounts of phonetic voicing in the fricative. The values of friction intensity, however, may also reflect other articulatory properties than just vocal fold vibrations. The lower values of friction intensity in items produced by the Norwegian speakers may possibly be a result of the different quality of the corresponding phoneme /ʊ/ in the Norwegian consonant inventory. This phoneme is usually described as labiodental approximant, due to its lack of friction (e.g. Davidsen-Nielsen, 1977; Kristoffersen, 2000: 25). In Czech, the corresponding phoneme is traditionally classified as a fricative, although recent research showed that in intervocalic positions, its acoustic properties (based on measures of harmonicity, duration and intensity profile) resemble more the characteristics of sonorants than obstruents (Skarnitzl and Volín, 2005). It seems plausible that the lower friction intensities in the fricative of both the non-native groups, as compared to natives,

Chapter 3

are caused by non-natives' tendency to use the phonetic qualities corresponding to the phonemes of their L1 inventories.

The speaking style effect on friction intensity in the word *of* is in accordance with the expected lower articulatory effort that may result in a more lenis consonant character in spontaneous as compared to read speech. In the word *to*, on the other hand, we revealed an opposite tendency to higher friction intensity in spontaneous speech, which seems difficult to relate to the expected tendencies to a higher degree of reduction in spontaneous speech.

4 Reduction of repeated mentions of content words

4.1 Introduction

As was presented in Section 1.2.1.3, a number of studies have focussed on the investigation of acoustic and phonetic parameters associated with the status of content words as new or given in a discourse. Previous research showed that repeated mentions of words within a discourse are shorter, less intelligible, and have lower F0 and more centralised vowel qualities. In addition, a greater degree of reduction of repeated mentions was found to occur in spontaneous speech and in communicative contexts. Most of the research on this topic, however, focussed on native production in English and Dutch, while only one study (Baker et al., 2011) confirmed similar reduction tendencies in non-native production.

The present study addresses the issue of repeated mention reduction comparing native productions (in Czech, English and Norwegian) and productions in non-native English, using task-elicited conversations. From previous research, it seems that this type of material is especially suitable for the investigation of the reduction of repeated mentions of content words. The investigation addresses durational, rhythmical and spectral aspects of the reduction of repeated mentions of content words. In addition, an attempt is made to control for discourse-related factors and prosodic structure.

4.2 Method

4.2.1 Speech material

The material used in this study consisted of spontaneous dialogues in Czech, English and Norwegian, elicited using a picture replication task (for details see Section 2.3.1). The English dialogues were produced by native British English speakers, as well as by two groups of non-native speakers of English, namely Czech and Norwegian speakers. More details about the material and the speakers can be found in Chapter 2.

4.2.2 Selected items

The lexical items investigated in this study were all nouns denoting objects discussed as part of the conversational task. As in the previous chapter, here we also aimed to select realisations of these words naturally integrated in surrounding speech. Since in spontaneous speech hesitations and other types of disfluencies are extremely frequent, it was not realistic to exclude all word tokens occurring in a close proximity of a disfluency. However, we excluded tokens where the investigated word itself contained a disfluency, mispronunciation or laughter, as well as words produced in isolation, regardless of the reason for the surrounding silence. Furthermore, we excluded tokens produced as part of the speaker's soliloquy⁴⁰. Lastly, we avoided using any lexical items that became an object of a misunderstanding in the course of the dialogue and required a clarification or additional explanation at some stage. It was assumed that items produced in such circumstances may have strongly dissimilar characteristics from neutral items produced during an unproblematic information exchange, and therefore, they should not be included in the studied sample. Moreover, articulation rate as well as several other factors related to discourse status and prosodic structure were taken into consideration. These factors will be discussed in more detail in Section 4.2.3.3.

We aimed to select up to four lexical items per speaker in each language-setting (note that the Czech and Norwegian speakers produced dialogues both in non-native English and in their native languages). For each lexical item, we used the speaker's first production of that word (first mention) and two later productions of the same word by the same speaker further in the dialogue (repeated mentions). The repeated mentions of a word were to be in the same form as the speaker's first mention of that word. This made the selection of suitable tokens particularly challenging in Czech and Norwegian. In Czech, nouns appear in a variety of forms due to case and grammatical number morphology, while Norwegian expresses the definite article of a noun by adding a

⁴⁰ Fairly easily distinguishable parts of a dialogue where the listening speaker talks as if to oneself while drawing, presumably to display understanding of the previous instructions and to indicate that he/she is busy and cannot process further instructions from the other speaker.

definite suffix to the word. Luckily, both in Czech and Norwegian some of the morphologically different word forms are still homophonic (e.g. the Norwegian words *gjerde* “fence” and *gjerdet* “the fence” are both pronounced as /jæ:re/). Moreover, in Czech it was necessary to pay attention to the lexical stress assignment. Although as a rule, stress in content words occurs on the first syllable of the word, a noun preceded by a monosyllabic preposition usually forms one stress group with the preposition where the stress is placed on the preposition (Palková, 1997: 280-282). Thus, we had to make sure not to include the cases where the observed noun does not form a stress group on its own. The three selected tokens per lexical item (spoken by the same speaker) will be referred to as a *triplet*. It should be noted that not all of the first productions of a word by a speaker were the first occurrences of the word in the dialogue. This will be discussed in more detail in Section 4.2.3.3. In repeated mentions of words, attention was paid to use only word tokens referring to the same object as the first mention of the word (e.g. if the first occurrence of the word *corner* was used in the context *corner of the room*, we would not include the token used in the context *corner of the bathtub*). The time elapsed since the word’s previous occurrence (by either of the speakers) was also taken into consideration. Here, we avoided using tokens that occurred after a period longer than 5 minutes since the last occurrence of the same word. A large majority of the repeated tokens, however, occurred within 2 minutes after the previous occurrence of the same word. Table 4.1 lists the numbers of triplets (lexical items in three repetitions produced by the same speaker) selected for each combination based on spoken language and the speakers’ L1. Each token was stored separately, including a context of approximately 2 seconds on each side.

Table 4.1: Numbers of triplets (lexical items in three repetitions produced by the same speaker) selected for each combination based on spoken language and speakers’ L1.

Language	Speakers' L1	Number of triplets
Czech	Czech	31
Norwegian	Norwegian	33
English	English	35
English	Czech	31
English	Norwegian	36

4.2.3 Acoustic analysis and additional observations

To prepare the selected items for automatic data extraction, a number of annotations were made using Praat (Boersma and Weenink, 2009). The annotations are described in Section 4.2.3.1. Based on the annotations, measurements were made and derived measures were calculated (see Section 4.2.3.2 for details). Section 4.2.3.3 describes some control variables relating to discourse status and certain aspects of prosodic structure in the speech fragments surrounding the selected tokens.

4.2.3.1 Annotations in content words

Figure 4.1 shows an example of annotations made in the content word tokens. Tier 1 (context) shows the orthographic transcription of the displayed speech fragment. On tier 2 (word), the observed content word is annotated in the second interval. In cases where the word is immediately followed by a silent pause, this pause is annotated in the third interval (p). On tier 3 (vowel), we annotated a stable portion within the vowel in the stressed syllable of the observed word, to be used for formant measurements (v). In polysyllabic words, we also annotated interval(s) corresponding to the part(s) of the word excluding the (primary) stressed syllable (unstressed). Determining the syllable boundaries was largely unproblematic. In a few cases where a consonant cluster occurred between the syllable nuclei, syllable boundaries were established using the maximal onset principle.

The criteria for placing boundaries between segments were taken from Machač and Skarnitzl (2009a). In spite of the difficulties with segmenting spontaneous speech, a good consistency could be achieved within the triplets (i.e. when segmenting the vowel or unstressed portion within the same word produced three times by the same speaker). In addition to the annotations described above, we annotated portions of clean speech (s), excluding silent pauses and easily separable disfluencies, on tier 4 (activity), and marked syllable nuclei as points on tier 5 (syllables). Since these annotations were meant to serve for calculation of the local articulation rate (see Section 4.2.3.2), they only stretch to approximately one second to each side from the observed word.

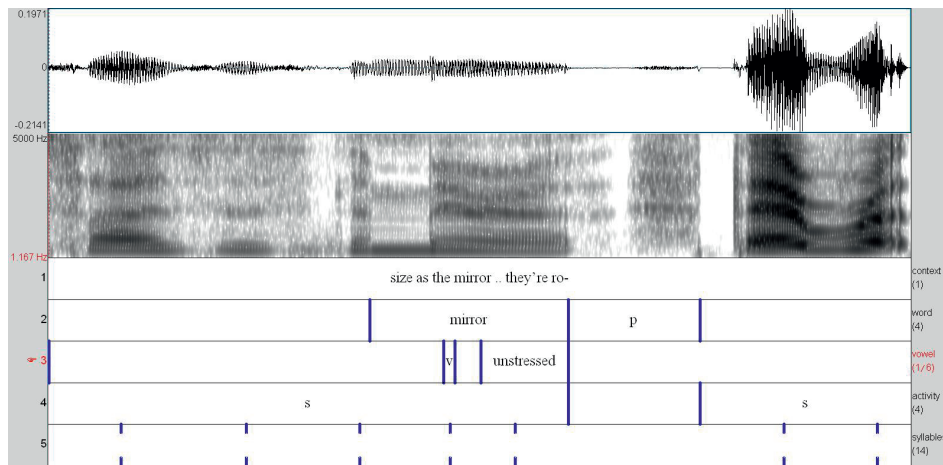


Figure 4.1: An example of annotations made in the observed content words (for details see text). **Tier 1** shows the orthographic transcription of the displayed speech fragment.

4.2.3.2 Acoustic measurements and calculating derived measures

The annotations of word boundaries and boundaries of unstressed portions in polysyllabic words allowed us to easily measure the whole word duration as well as the stressed syllable duration in milliseconds (note that in monosyllabic words, the values of these two variables are equal). However, the different subsets (defined by the combination of spoken language and speakers' L1) within the dataset contained different lexical items, and therefore, their mean word durations were inevitably different. In order to reduce the variance due to the different lexical items involved, we inspected some additional durational measures, including the mean syllable duration (whole word duration divided by the number of syllables) and the mean unstressed syllable duration (duration of unstressed portion⁴¹ divided by the number of unstressed syllables⁴²). The latter measure was, naturally, only relevant in polysyllabic words. Lastly, in order to capture the variability in durations of stressed and unstressed

⁴¹ This included all syllables except for the syllable with primary stress.

⁴² All syllables that did not have primary stress

Chapter 4

syllables in polysyllabic words, we calculated a *ratio of unstressed-to-stressed syllable duration* using a formula:

$$\text{ratio}_{\text{uns:str}} = \text{dur}_{\text{uns}} / \text{dur}_{\text{str}},$$

where dur_{uns} is the mean duration of unstressed syllable and dur_{str} is the duration of the syllable with the primary stress. A similar approach has been used in a number of studies on lexical stress and rhythm in non-native production. For example, Lee et al. (2006) used a ratio of durations of the unstressed to the stressed vowel in a given word to investigate the realisation of English lexical stress by Japanese and Korean learners. Volín (2005) calculated a durational reduction coefficient as a ratio of the duration of the stressed vowel to the mean duration of the unstressed vowels within the word. This measure was applied in a small set of polysyllabic English words produced by natives, and Czech speakers judged as having a strong non-native accent. Furthermore, Trofimovich and Baker (2006) defined a measure of stress-timing as the ratio of mean unstressed syllable duration to mean stressed syllable duration in a study on acquisition of L2 suprasegmentals. Although using vowel durations to calculate ratios would possibly offer more reliable results than using durations of whole syllables, we preferred to avoid the difficulties associated with the detailed segmentation of spontaneous speech signal and only measured the syllable durations. As a result of that, the values of unstressed-to-stressed syllable duration ratio in particular word tokens reflect not only the amount of stress-induced durational contrast that we desire to measure, but also the complexity of the syllables occurring in the given words.

Apart from the durational measures, we intended to inspect vowel quality in the stressed syllable (in tokens containing monophthongs in their stressed syllable). The mean formant values (F1 and F2) in Bark were measured in the stable portion of the vowel, as annotated in the textgrid. To eliminate errors in automatic formant tracking, an additional semi-automatic method was used to detect any abrupt jumps between nearby values of the same formant, as determined by the Praat formant-tracking algorithm. The items where a jump of over 3 Bark within 12.5 ms was present in the formant

measurements were then manually checked, and the automatically obtained values were corrected to match the formants, as determined by a careful inspection of the spectrogram. The resulting formant values were used to calculate a measure of vowel spectral contrast, expressing a one-dimensional distance of a vowel to the vowel-space centre. To be able to calculate this value, we first needed to determine a point that would serve as a vowel-space centre (referred to as *centroid*). We calculated female and male centroids separately for each of the observed languages (using formant values reported by Hedbávná, 2004, for Czech; Deterding, 1997, for English and van Dommelen, 2011, for Norwegian). Each centroid was determined as a point in the formant space representing the mean F1 and mean F2 value of all vowel phonemes in the inventory of the respective language in Bark. Table 4.2 lists the resulting formant values of female and male centroids in each language. Although the differences between centroids in the three languages are not very large, using separate centroids for the three languages was considered more appropriate with regard to presumable differences in language-specific articulatory settings (cf. Gick et al., 2004; Wilson, 2006). The *distance to the centroid* was then calculated as the Euclidean distance in the F1-F2 space between the given vowel and the gender-specific centroid for the given language in Bark.

Table 4.2: Formant values (in Bark) of female and male centroids in Czech (CZE), British English (ENG) and Norwegian (NOR).

		F1 (Bark)	F2 (Bark)
CZE	females	5.12	11.05
	males	4.47	10.22
ENG	females	5.66	11.43
	males	4.67	10.41
NOR	females	5.21	11.51
	males	4.40	10.44

It needs to be mentioned that the measure of distance to the centroid was calculated only for tokens containing monophthongs in their stressed syllable (i.e. not words that had a diphthong or a syllabic /r/ as nucleus of their stressed syllable). Moreover, in a few

Chapter 4

cases where vowel formant measurements were unreliable due to disturbing background noise (e.g. overlapping speech), the values for all triplet members were excluded from analyses. The same was done whenever a vowel's distance to the centroid was less than 1 Bark, assuming that the distance to the centroid can serve as a reliable measure of centralisation particularly in peripheral vowels. In addition, we checked all items where the measured vowel occurred in the neighbourhood of a nasal consonant. For items where the appropriate placement of formants corresponding to the vowel quality could not be reliably determined due to a strong influence of nasalisation, the values for all triplet members were excluded from the analyses.

A similar measure of distance to the centroid was used by Koopmans-van Beinum (1980: 55-62) to investigate the differences in vowel quality in different speaking styles in Dutch. Harmegnies and Poch-Olivé (1992) calculated the Euclidean distances of vowels from schwa (defined as: $F1 = 500$ Hz, $F2 = 1500$ Hz) and the centralisation indices as the differences between vowel-schwa distance for a given vowel in laboratory speech, and in spontaneous speech in Spanish. Similarly, Laan (1997) calculated the Euclidean distances of vowels in different speaking styles from "ideal" formant frequencies measured in vowels produced in isolation. A slight drawback of our approach may be the use of a common centroid for all female / male speakers rather than using speaker-specific centroids. Unfortunately, the available recordings (both spontaneous speech and read text; for details see Chapter 2) were not specifically designed for investigations of speakers' vowel systems, and any efforts to collect representative sample from these recordings were expected to provide rather unreliable results. We assumed, however, that the possible inaccuracies in centroid position were not very large, and this simplification should be appropriate for the purpose of *comparing* distances between the successive mentions of words, rather than using their absolute values to compare between speakers. Moreover, as mentioned above, all vowels with a distance to the centroid less than 1 Bark, which were at most risk to present misleading values of distance to the centroid, were excluded from the analyses.

Apart from the measures directly concerning the observed word tokens, we calculated the local articulation rate in the speech fragment surrounding the observed word (approximately 1 second on each side). The value was determined as the number of syllables divided by clean speech duration in seconds, in the speech fragment *excluding* the observed content word. The procedure was comparable⁴³ to that used in the previous chapter investigating the realisations of English function words (see Section 3.2.3.4 for details).

4.2.3.3 Control factors: discourse status and prosodic structure

Previous research on reduction in speech has shown an undeniable influence of a number of factors on the degree of reduction. While we attempted to control for some of them by excluding tokens with deviating properties (see Section 4.2.2), other influential factors still remain to be taken into consideration. We assumed that factors related with word predictability in a given discourse context and certain aspects of prosodic structure in the speech fragments surrounding the selected tokens would be the most relevant.

Among the many variables that could describe different aspects of how a word token relates to the ongoing discourse, we focussed on whether the first triplet member was the first occurrence of the word within the dialogue. As was mentioned previously, in some cases the first speaker's production of a word was already preceded by the production of the same word by the other speaker. This was mainly due to the unequal roles of the two speakers in the conversational task they carried out. A binary variable *dialogue firstness* coded for each triplet, whether its first member was the first occurrence of the word within the dialogue (1) or whether the word had been produced previously by the other speaker (0).

⁴³ In the previous chapter, local articulation rate was calculated using the whole speech fragment, including the investigated function word. It was not considered necessary to exclude function words from the local articulation rate calculation since they constituted only a negligible part of the speech fragments.

Chapter 4

Regarding prosodic structure, we considered two elements that may play a major role in word realisations: prosodic prominence (accentuation) and position relative to prosodic boundaries. To evaluate the degree of accentuation we used the judgements of three listeners. The listeners were phoneticians trained to analyse various aspects of speech signals. They were, however, not particularly experienced with prosodic annotations. In addition, all listeners were either native speakers or very proficient speakers of the language, the tokens of which they evaluated. The listeners were presented with speech fragments containing a particular word with a context of approximately 2 seconds on each side and they were asked to evaluate the perceived degree of accentuation on the observed words. The instructions provided to the listeners described accentuation as prosodic prominence characterised by a pitch movement associated with a word's stressed syllable, as well as by increased intensity and lengthening (cf. Terken and Hirschberg, 1994). Moreover, nuclear accent was defined as the main prominence within an intonational phrase. The degree of prominence in deaccented words, on the other hand, was compared to that of function words. The listeners were encouraged to rely on their intuitive perceptual impression of prominence rather than to focus on individual suprasegmental features. Within each listener's evaluation, the tokens evaluated as carrying an accent or nuclear accent were later pooled together, as one category (accented). Using the listeners' evaluations, the word tokens that were evaluated as accented by two or more out of the three judges were coded as *accented*. The agreement between the listeners' evaluations was on average quite high. Pairwise agreement was on average 95% for the English materials (including both native and non-native items), 89% for the Norwegian materials and 90% for the Czech materials. The Kappa value used to determine the reliability of agreement between evaluations (e.g. Poesio and Vieira, 1998) was found to be 0.9 for the English subset of the data, 0.77 for the Norwegian part and 0.8 for the Czech part of the material. The Kappa statistic $K > 0.8$ is generally considered to indicate good reliability, whereas $0.8 > K > 0.68$ allows drawing tentative conclusions (Carletta, 1996; cited in Poesio and Vieira, 1998).

In addition, to control for the effects of final lengthening we coded all cases where the observed token was followed by a major prosodic boundary. For that purpose, we measured the duration of a silent pause immediately following the word (if present). All cases where the pause duration exceeded 150 ms were then coded as *final*. In a number of cases, there was no silent pause present in the signal, but other cues indicated a strong prosodic break. Thus, we also coded a token as *final*, whenever the prosodic disjuncture after the word corresponded to ToBI break index 4 (describing a full intonational phrase boundary; based on definitions in Beckman and Ayers Elam, 1997). The remaining tokens were coded as *non-final*. Although previous research has shown that final lengthening may also affect units followed by prosodic boundaries of lower degrees, we preferred to keep the analyses simple by choosing more robust criteria. The present material consists of unprepared spontaneous dialogues, partly produced by less fluent non-native speakers, and even a simple evaluation of prosodic boundaries therefore presents a considerable challenge. The reliability of any more fine-grained analyses was thus considered doubtful.

4.2.4 Praat scripts

As in Chapter 3, the data were obtained from the prepared textgrids using Praat scripts. This procedure also allowed us to check the consistency of annotations in several ways. The main output consisted of tables (text files) containing all the data ready for use in statistic software.

4.2.5 Statistical analyses

Statistical analyses were carried out using the same tools and conventions as described in Section 3.2.5. All analyses were run separately on a subset of *L1 data* (items produced by native speakers of Czech, Norwegian and English) and a subset of *English data* (items in English produced by native speakers and Czech and Norwegian learners). In this chapter, we mainly used repeated-measures ANOVA with repeated mention (also referred to as *repetition*) as the within-group factor and the grouping based on actual spoken language or the speakers' L1 as the between-group factor. The aim was to compare the three different languages (Czech, English and Norwegian) spoken by

native speakers, and to compare the productions in English by native speakers and two groups of non-native speakers. Moreover, to be able to determine the effect of some control variables on the observed measures, mixed-effects models were used, using the *lme4* package for R (R Development Core Team, 2010).

4.3 Results

The following sections present the results describing durational, rhythmical and spectral aspects of the reduction of repeated mentions of content words. As mentioned above, all analyses are carried out separately on a subset of *L1 data* (items produced by native speakers of Czech, Norwegian and English) and a subset of *English data* (items in English produced by native speakers and Czech and Norwegian learners). Section 4.3.1 presents the results for the main effects of repeated mention and speaker group, comparing the three different languages, as well as English spoken by native and non-native speakers. Section 4.3.2 focusses on describing the effects and interactions of the discourse-related control factor *dialogue firstness*. Finally, Section 4.3.3 deals with the control factors relating to prosodic structure (*accentuation* and *finality*).

4.3.1 Effect of repeated mention and speaker group

4.3.1.1 Articulation rate and durational measures

As in the previous chapter dealing with function words, we assumed that local articulation rate⁴⁴ in the speech fragment surrounding the observed words may contribute to the durational variation in the dataset. In this case, however, we could not determine the correlation of articulation rate with word durations easily, due to the heterogeneous composition of the dataset. The sample contained very different words, such as *tap* vs. *harmonica*. Therefore, we investigated the correlation of articulation rate

⁴⁴ Calculated as the number of syllables divided by clean speech duration in seconds

with the duration of stressed syllable in the word, and with the mean duration of syllable within the word (calculated as the word duration divided by the number of syllables in the word) instead. However, it needs to be noted that although these measures of duration were more appropriate to reduce variability due to different numbers of syllables in the different lexical items within the dataset, they still did not dispose of the variance due to factors such as different syllable complexities, presence of phonologically long vs. short vowels, etc. In addition, the articulation rate data were logarithmically transformed to improve their linear relationship with the observed durational variables. The results showed that both the correlation of stressed syllable duration and the correlation of mean syllable duration with logarithmically transformed articulation rate were highly significant, but achieved weak negative correlation coefficients ($r = -0.275$; $N = 492$; $p < 0.001$ and $r = -0.302$; $N = 492$; $p < 0.001$, for the two measures, respectively).

Further, we inspected the effects of repeated mention and speaker group on articulation rate. In the subset of *L1 data*, a repeated-measures ANOVA with repeated mention as the within-group factor and spoken language as the between-group factor showed only a significant effect of spoken language ($F(2, 94) = 9.25$; $p < 0.001$), while neither the effect of repeated mention nor its interaction with spoken language reached significance ($F(2, 188) < 1$ and $F(4, 188) < 1$, respectively). Likewise, in a comparable analysis in the subset of *English data*, only a significant effect of spoken language ($F(2, 95) = 6.76$; $p = 0.002$) was revealed. Neither the effect of repeated mention nor its interaction with spoken language reached significance ($F(2, 190) = 2.17$; $p = 0.117$ and $F(4, 190) < 1$, respectively). These results suggest that there are no overall significant differences in articulation rates across the repeated mentions of the observed words. Therefore, we may assume that any effects of repeated mention revealed in upcoming analyses are *not* due to incidental differences in articulation rate. On the other hand, we revealed significantly lower articulation rates in items spoken by native English speakers (5.24 syll/s), compared to articulation rates of Czech and Norwegian speakers in their native languages (6.01 syll/s and 6.35 syll/s, respectively). The native English speakers' articulation rate was, however, still significantly higher than articulation rates of English

Chapter 4

spoken by both non-native speaker groups (4.22 syll/s and 4.39 syll/s for Czech and Norwegian speakers, respectively).

The lack of articulation rate differences across the repeated mentions of words, as well as the relatively weak correlation of articulation rates with durational measures, revealed by the analyses presented above, indicate that there may be no need to include articulation rate in the analyses of the main experimental factors. A possibility to use ANCOVA does not seem reasonable due to the revealed dependence of articulation rate on speaker group. Including articulation rate as a (time-varying) covariate in the main analyses of the effects of repeated mentions and speaker group may then result in assigning the variance due to speaker group to articulation rate variability. Moreover, due to the heterogeneity of the dataset, it was not possible to use any simple method to normalise word durations relative to the articulation rate in a comparable way as in Chapter 3 on the reduction of English function words (cf. Section 3.3.1.1). Therefore, despite the effect of articulation rate on durational variables, we chose to disregard this factor in the analyses of the main effects of repeated mention and speaker group, as well as in further analyses using repeated-measures ANOVA. Still, articulation rate will be used as an additional predictor in the analyses of effects of the control factor *finality* carried out using mixed-effect models (see Section 4.3.3).

4.3.1.2 Durational reduction of repeated mentions

As suggested by previous research, repeated mentions of content words may be expected to show durational reduction. Due to the heterogeneous nature of the inspected sample, which consisted of triplets of repeated mentions of different lexical items, we decided to use two different measures relating to duration. Obviously, whole word duration was expected to differ as a result of different numbers of syllables in the observed words, as well as of different syllable complexities. While neither of these differences matters when comparing the repeated mentions of a particular word (using a repeated-measures ANOVA), words of very different inherent duration may not be expected to show a uniform durational reduction. This may also complicate the

interpretation of interactions between the repeated mention and speaker group. In order to compensate for the variability due to different numbers of syllables in the represented words, we inspected another durational measure: mean syllable duration. Even this measure, however, cannot be expected to dispose of the variance due to syllable structure differences or the presence and realisation of lexical stress. Therefore, the results have to be interpreted with caution, and neither of the measures should be assumed to provide an unambiguous answer. Additional variables that provide more detail about the temporal patterns resulting from rhythmical regularities of the observed languages will be analysed in Section 4.3.1.2.

The first measure that we inspected was the whole word duration. Figure 4.2 shows the mean durations of the three successive mentions of words, comparing the three languages spoken by their native speakers, and Figure 4.3 displays the mean word durations in repeated mentions of English words spoken by natives and by Czech and Norwegian speakers. The connecting lines in these figures (and other figures throughout this chapter) are not meant as interpolation (between the repeated mentions of words), but serve for a visualisation of the interaction between the word mention and speaker groups.

As for the subset of *L1 data* (Figure 4.2), it seemed that the duration of words decreases in repeated mentions. Pooled across the groups based on spoken language, the duration of the first mention was 470 ms, while second and third mentions were considerably shorter (416 ms and 403 ms, respectively). We may also observe that while the duration of Czech and English items decrease only when comparing the first with later mentions, Norwegian speakers seem to reduce the duration further in later repetitions of the same word. More detailed statistical information relating to word durations, as well as to other variables inspected in this chapter, are listed in Appendix C.

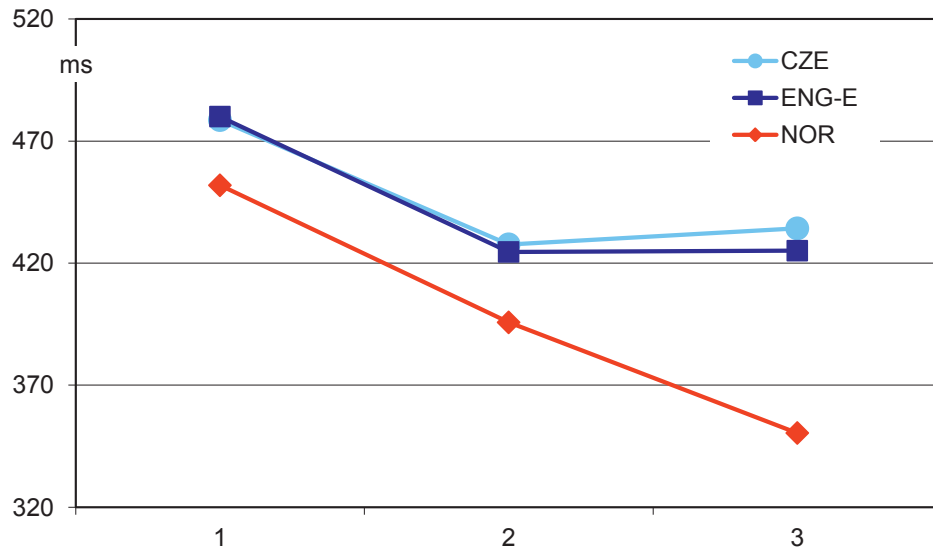


Figure 4.2: Mean durations (in ms) of the first mention and two repeated mentions of content words in Czech (CZE), English (ENG-E) and Norwegian (NOR).

A repeated-measures ANOVA with repeated mention as the within-group factor and spoken language as the between-group factor showed a significant effect of repeated mention ($F(2, 192) = 12.0$; $p < 0.001$) but neither the spoken language nor its interaction with repeated mention reached significance ($F(2, 96) = 1.40$; $p = 0.253$ and $F(4, 192) < 1$, respectively). We may speculate that the interaction did not reach significance due to a large dispersion of the observed values. The tests of within-subjects contrasts revealed that the duration of both the second and third successive mentions of words differed significantly from the first mention ($F(1, 96) = 12.8$; $p = 0.001$ and $F(1, 96) = 20.6$; $p < 0.001$, respectively).

Similarly, in the subset of *English data* (Figure 4.3), durations of repeated mentions of words were shorter than their first mentions (pooled across the speaker groups 498 ms for the first mention and 431 ms and 424 ms for the two repeated mentions). Durations of first mentions produced by Czech speakers seemed somewhat longer compared to first mentions of items spoken by the other two groups while the mean durations of

subsequent mentions of the words were very similar across the speaker groups. A repeated-measures ANOVA with repeated mention as the within-group factor and speakers' native language as the between-group factor showed a significant effect of repeated mention ($F(2, 198) = 17.6$; $p < 0.001$) but neither the speakers' L1 nor its interaction with repeated mention reached significance ($F(2, 99) < 1$ and $F(4, 198) = 1.42$; $p = 0.229$, respectively). Again, the lack of significant interaction may be possibly attributed to a large dispersion of the durational values. The tests of within-subjects contrasts showed that the duration of both the second and third successive mentions of words differed significantly from the first mention ($F(1, 99) = 22.3$; $p < 0.001$ and $F(1, 99) = 25.7$; $p < 0.001$, respectively).

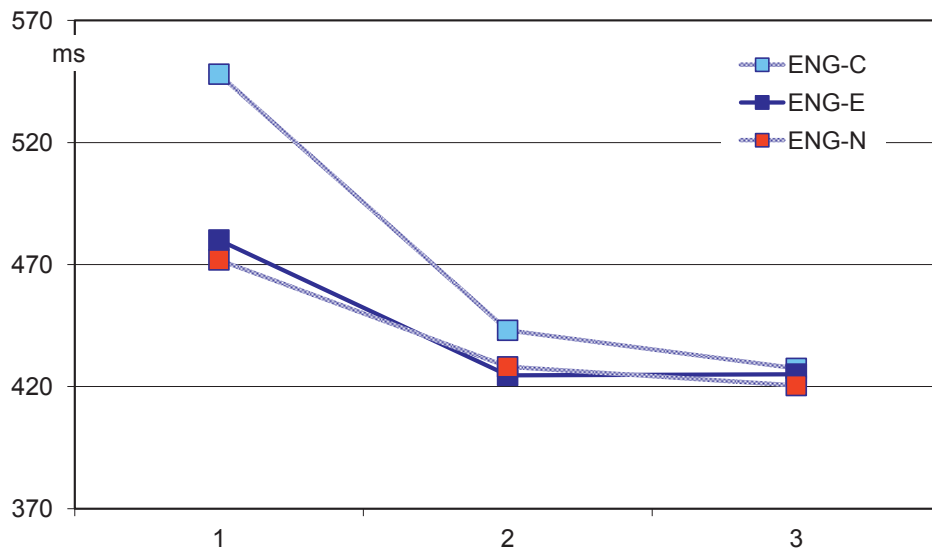


Figure 4.3: Mean durations (in ms) of the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).

As we mentioned previously, the measure of whole word duration does not control for the properties of the actual lexical items represented in the samples, and therefore, it may present distorted results as to between-group factors and interactions. One of the

Chapter 4

potentially most influential properties is the number of syllables in the observed lexical items. In our data, the samples for different groups based on spoken language differed substantially: while in Czech and Norwegian, there was only a small proportion of monosyllabic words (23% for Czech and 15% for Norwegian), monosyllabic words made up more than half of the items in the samples of English (58%, 60% and 67% in samples of English words spoken by Czech, native English and Norwegian speakers, respectively). The actual numbers of monosyllabic and polysyllabic items (triplets) in each sample are shown in Table 4.3. It is possible that this strong imbalance in the composition of different group samples obscures the between-group differences or interactions.

A possible way to reduce the bias due to the different number of syllables in the samples of lexical items representing the different groups based on spoken language and speakers' L1 was using the measure of mean syllable duration (see Figures 4.4 and 4.5). As mentioned above, we should be aware that even this measure does not dispose of the variance due to different syllable complexities, presence of phonologically long vs. short vowels or the difference between stressed and unstressed syllables.

Table 4.3: Numbers and percentages of monosyllabic and polysyllabic items (triplets) in each group based on spoken language and speakers' L1.

	all items	monosyllabic items	monosyll. (%)	polysyllabic items	polysyll. (%)
CZE	31	7	23	24	77
ENG-C	31	18	58	13	42
ENG-E	35	21	60	14	40
ENG-N	36	24	67	12	33
NOR	33	5	15	28	85
Total	166	75	45	91	55

As can be observed in Figure 4.4, apart from a general slight decrease in duration over the repeated mentions, there is a large effect of spoken language. Pooled across the word repetitions, mean syllable duration in English words (produced by natives) was much higher (332 ms) than in Czech (224 ms) and Norwegian (209 ms). This large

difference may be the result of a large proportion of monosyllabic words in the English samples (see Table 4.3 above), as well as of higher syllable complexity of typical English (stressed) syllables. On the other hand, the mean syllable durations in English words produced by speakers with different L1s were rather similar (see Figure 4.5). Pooled across the word repetitions, the mean syllable durations were 359 ms, 332 ms and 348 ms for Czech, English and Norwegian speakers, respectively.

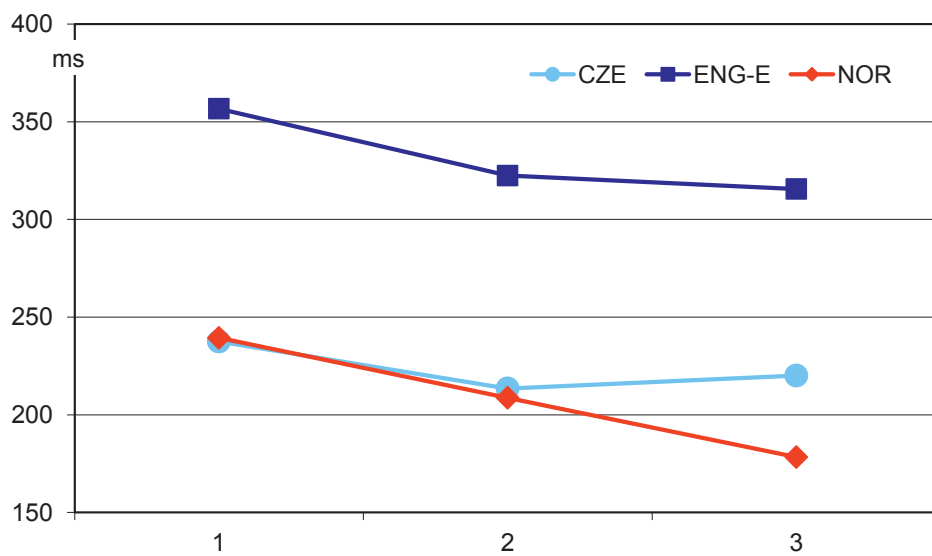


Figure 4.4: Mean syllable durations (in ms) in the first mention and two repeated mentions of content words in Czech (CZE), English (ENG-E) and Norwegian (NOR).

In the subset of *L1 data* (Figure 4.4), a repeated-measures ANOVA with repeated mention as the within-group factor and spoken language as the between-group factor using a Huynh-Feldt correction⁴⁵ showed a significant effect of repeated mention ($F(1.938, 186.075) = 7.75$; $p = 0.001$) but no significant interaction ($F(3.877, 186.075) < 1$). The tests of within-subjects contrasts showed that the mean syllable duration in both

⁴⁵ Mauchly's test of sphericity indicated that the sphericity assumption had been violated, $\chi(2) = 7.24$; $p = 0.027$, and therefore, a Huynh-Feldt correction was used.

Chapter 4

the second and third successive mentions of words differed significantly from the first mention ($F(1, 96)= 7.89$; $p= 0.006$ and $F(1, 96)= 11.7$; $p= 0.001$, respectively). Since Levene's test indicated unequal variances with high significance ($p< 0.001$), we performed the test of between-subject effects using non-parametric tests. The Kruskal-Wallis test showed a significant effect of spoken language ($\chi^2(2)= 59.2$; $p< 0.001$). A post-hoc test using Mann-Whitney tests with Bonferroni correction showed that native English speakers' mean syllable duration differed significantly from both Czech speakers' ($p< 0.001$) and Norwegian speakers' ($p< 0.001$) values.

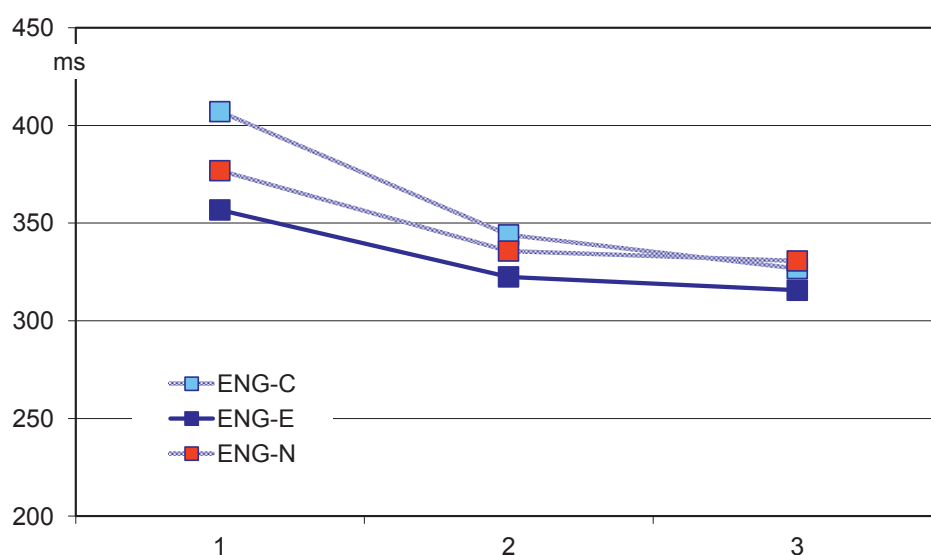


Figure 4.5: Mean syllable durations (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).

In the subset of *English data* (Figure 4.5), a repeated-measures ANOVA with repeated mention as the within-group factor and speakers' L1 as the between-group factor showed a significant effect of repeated mention ($F(2, 198)= 12.1$; $p< 0.001$) but neither the speakers' native language nor its interaction with repeated mention reached significance ($F(2, 99)< 1$ and $F(4, 198)< 1$, respectively). The tests of within-subjects contrasts showed that the mean syllable duration in both the second and third successive

mentions of words differed significantly from the first mention ($F(1, 99) = 14.9$; $p < 0.001$ and $F(1, 99) = 18.3$; $p < 0.001$, respectively).

4.3.1.3 Rhythmical aspects of reduction of repeated mentions

In order to describe the temporal structure of the observed words in more detail, we also inspected the stressed syllable duration and the mean unstressed syllable duration (in polysyllabic words). Although all words within the dataset have a syllable with primary stress, it was assumed that stressed syllable characteristics differ depending on whether the word is monosyllabic or polysyllabic (cf. Lehiste, 1972), and therefore, it may be more appropriate to describe stressed syllable durations separately for subsets of monosyllabic and polysyllabic words. To enable a more systematic description of the results, we will refer to both of these as analyses of *stressed syllable durations* in the following text. It has to be noted, however, that in the case of monosyllabic words, this measure actually represents whole word durations. Figure 4.6 shows the mean stressed syllable durations in monosyllabic words, comparing the three languages spoken by their native speakers. It needs to be mentioned that in this comparison, the groups of items in Czech and Norwegian contained only a very few observations, so the relations in the figure may not be representative. Figure 4.7 then shows the mean stressed syllable durations in monosyllabic English words spoken by natives and by Czech and Norwegian speakers.

Figure 4.6 shows considerable differences of the stressed syllable durations in monosyllabic words between the three languages (*L1 data* subset). Pooled across the three mentions of words, mean stressed syllable duration (in monosyllabic words) was 278 ms in Czech, 398 ms in English and 353 ms in Norwegian. The decrease in duration in the repeated mentions of words was, apart from the Norwegian items, rather slight. A repeated-measures ANOVA with repeated mention as the within-group factor and spoken language as the between-group factor showed only a marginally significant effect of repeated mention ($F(2, 60) = 3.11$; $p = 0.052$) and a significant effect of spoken language ($F(2, 30) = 4.71$; $p = 0.017$), while surprisingly there was no significant

Chapter 4

interaction ($F(4, 60) = 1.13$; $p = 0.350$). The tests of within-subjects contrasts showed that only the stressed syllable duration in the third mention of monosyllabic words differed significantly from the first mention ($F(1, 30) = 4.80$; $p = 0.036$). Tukey's post-hoc test showed that the spoken language effect was due to native English speakers' stressed syllable duration differing significantly from Czech speakers' values ($p = 0.013$). The absence of a significant interaction in spite of the sharply dissimilar patterns of monosyllabic words' shortening between the speaker groups (note the data for Norwegian speakers in Figure 4.6) may be explained by the very low number of observations in some of the groups (i.e. only 7 triplets for Czech speaker group and 5 triplets for Norwegians).

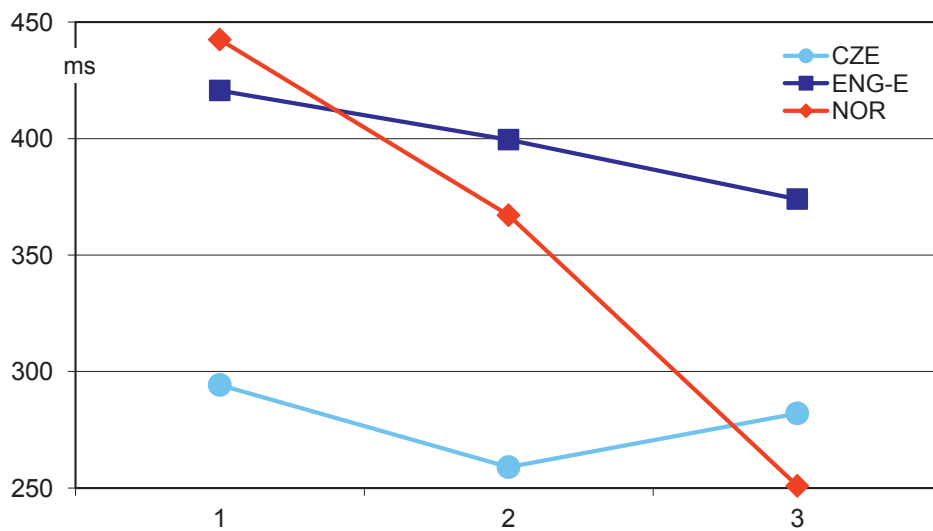


Figure 4.6: Mean stressed syllable durations (in ms) in the first mention and two repeated mentions of content words in monosyllabic words in Czech (CZE), English (ENG-E) and Norwegian (NOR).

The stressed syllable durations in monosyllabic words for the subset of *English data* are displayed in Figure 4.7. Here we can observe a moderate decrease in stressed syllable duration over the repeated mentions as well as small differences between the durations of items produced by speakers with different L1s. Pooled across the groups based on speakers' L1, the stressed syllable duration in the first mention was 455 ms, while

durations in second and third mentions were considerably shorter (413 ms and 391 ms, respectively). A repeated-measures ANOVA with repeated mention as the within-group factor and speakers' native language as the between-group factor showed a significant effect of repeated mention ($F(2, 120) = 5.91$; $p = 0.004$) but neither the speakers' L1 nor its interaction with the repeated mention reached significance ($F(2, 60) = 1.13$; $p = 0.329$ and $F(4, 120) < 1$, respectively). The tests of within-subjects contrasts showed that the stressed syllable duration in both the second and third successive mentions of monosyllabic words differed significantly from the first mention ($F(1, 60) = 4.91$; $p = 0.030$ and $F(1, 60) = 9.76$; $p = 0.003$, respectively).

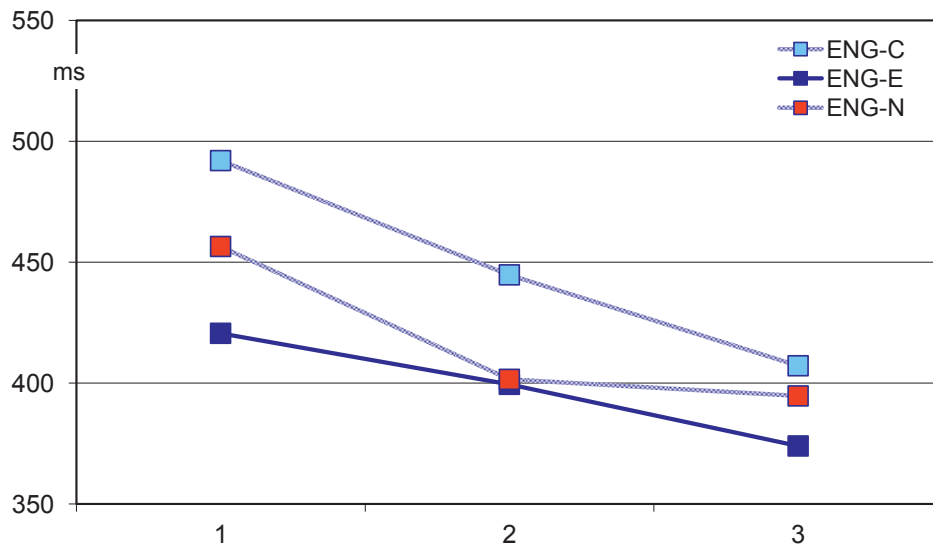


Figure 4.7: Mean stressed syllable durations (in ms) in the first mention and two repeated mentions of content words in the subset of monosyllabic words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).

Next, we inspected the stressed syllable durations for the subset of polysyllabic words. Overall, stressed syllable durations seemed to be considerably longer in monosyllabic words than in polysyllabic words (cf. Figures 4.6 and 4.7 vs. Figures 4.8 and 4.9). This may be explained by processes such as polysyllabic word compression, or final lengthening affecting the stressed syllable in monosyllabic words (as compared to

Chapter 4

polysyllabic words, where the last syllable, whether stressed or unstressed would be the domain of final lengthening). Figure 4.8 shows the mean stressed syllable durations in polysyllabic words, comparing the three languages spoken by their native speakers, and Figure 4.9 shows the mean stressed syllable durations in polysyllabic English words spoken by natives, and by Czech and Norwegian speakers. In the *L1 data* subset, we could observe moderate differences between the three languages. Pooled across the three mentions of words, mean stressed syllable duration (in polysyllabic words) was 181 ms in Czech, 216 ms in English and 191 ms in Norwegian. In addition, stressed syllable duration decreases in repeated mentions compared to first mentions of words. Pooled across the groups based on spoken language, the stressed syllable duration in the first mention was 213 ms, while durations in second and third mentions were shorter (182 ms and 183 ms, respectively).

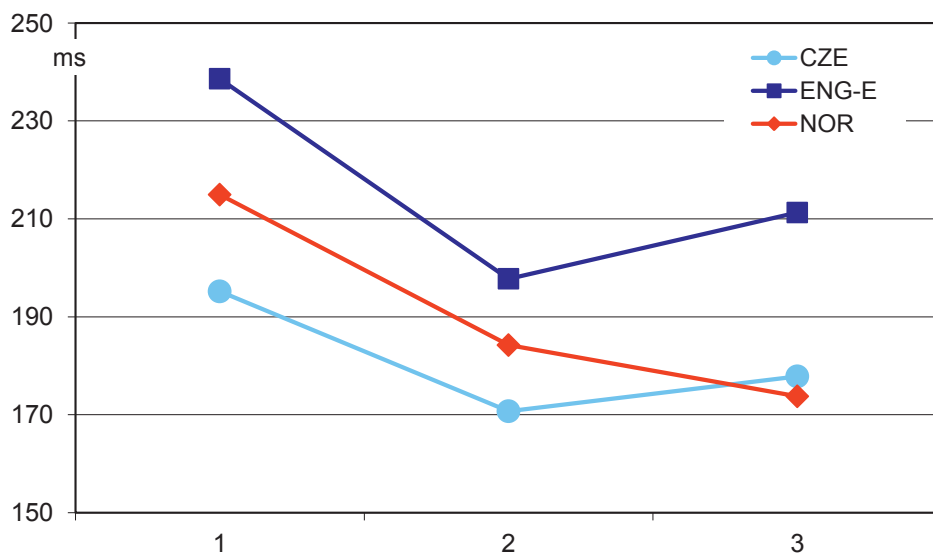


Figure 4.8: Mean stressed syllable durations (in ms) in the first mention and two repeated mentions of content words in polysyllabic words in Czech (CZE), English (ENG-E) and Norwegian (NOR).

A repeated-measures ANOVA with repeated mention as the within-group factor and spoken language as the between-group factor showed a significant effect of repeated

mention ($F(2, 126) = 13.0$; $p < 0.001$), but neither the effect of spoken language nor the interaction reached significance ($F(2, 63) = 1.42$; $p = 0.248$ and $F(4, 126) < 1$, respectively). The fact that the main effect of spoken language did not reach significance is presumably due to the large within-group variance. The tests of within-subjects contrasts showed that the stressed syllable duration in both the second and third successive mentions of polysyllabic words differed significantly from the first mention ($F(1, 63) = 18.9$; $p < 0.001$ and $F(1, 63) = 20.7$; $p < 0.001$, respectively).

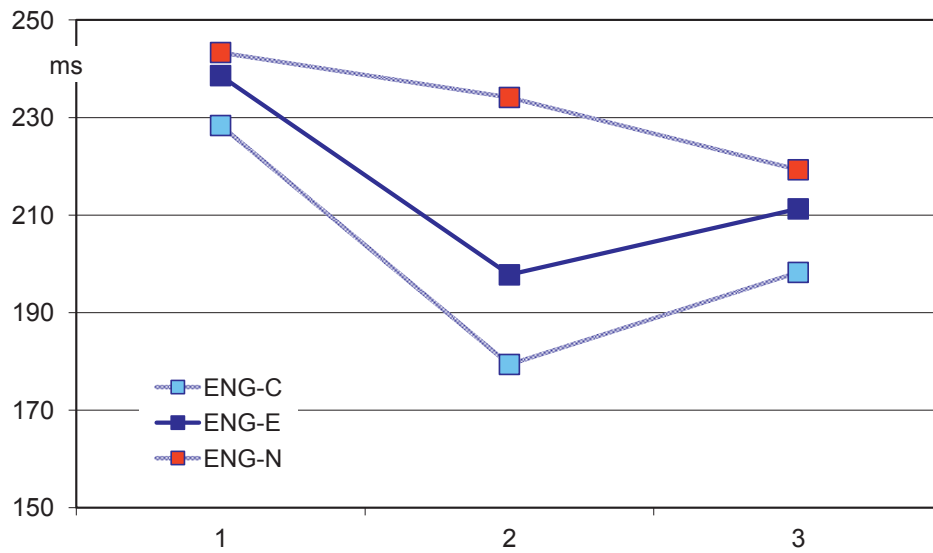


Figure 4.9: Mean stressed syllable durations (in ms) in the first mention and two repeated mentions of content words in the subset of polysyllabic words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).

In the *English data* subset, stressed syllable durations (in polysyllabic words) were also substantially shorter in repeated mentions compared to first mentions of words, but the differences between groups of items produced by speakers with different L1s were less obvious (see Figure 4.9). A repeated-measures ANOVA with repeated mention as the

Chapter 4

within-group factor and speakers' native language as the between-group factor using a Huynh-Feldt correction⁴⁶ showed a significant effect of repeated mention ($F(1.698, 61.145) = 9.03$; $p = 0.001$) but neither the speakers' L1 nor its interaction with repeated mention reached significance ($F(2, 36) < 1$ and $F(3.397, 61.145) = 1.21$; $p = 0.317$, respectively). The tests of within-subjects contrasts showed that the stressed syllable duration in both the second and third successive mentions of polysyllabic words differed significantly from the first mention ($F(1, 36) = 13.8$; $p = 0.001$ and $F(1, 36) = 7.81$; $p = 0.008$, respectively).

Further, in order to describe how the syllable with the primary word stress and the remaining portion of the word (referred to as the *unstressed portion*, even though in some cases there may be a secondary stress present at one of the syllables) contribute to the overall duration of polysyllabic words, we inspected two more measures (in the subset of the data containing only the polysyllabic words). The first of these measures was unstressed syllable duration. The mean unstressed syllable durations comparing the three languages spoken by their native speakers are displayed in Figure 4.10. Figure 4.11 shows the mean unstressed syllable durations in English words spoken by natives and by Czech and Norwegian speakers. In the *L1 data* subset, the mean unstressed syllable durations in repeated mentions were slightly shorter than in the words' first mentions. Pooled across the observed languages, the durations were 229 ms in first mentions and 203 ms and 200 ms in the two repeated mentions. There also seemed to be differences in unstressed syllable durations between the groups of items in the three different languages. Pooled across the three mentions of words, mean duration in Czech items was 230 ms, in English items it was 245 ms and in Norwegian items it was 177 ms. A repeated-measures ANOVA with repeated mention as the within-group factor and spoken language as the between-group factor showed a significant effect of repetition ($F(2, 126) = 5.22$; $p = 0.007$) and spoken language ($F(2, 63) = 4.06$; $p = 0.022$), while the interaction did not reach significance ($F(4, 126) = 1.37$; $p = 0.249$). The tests of within-

⁴⁶ Mauchly's test of sphericity indicated that the sphericity assumption had been violated, $\chi(2) = 11.9$; $p = 0.003$, and therefore, a Huynh-Feldt correction was used.

subjects contrasts showed that the unstressed syllable duration in both the second and third successive mentions of words differed significantly from the first mention ($F(1, 63) = 6.94$; $p = 0.011$ and $F(1, 63) = 8.93$; $p = 0.004$, respectively). Tukey's post hoc test showed that the spoken language effect was due to native English speakers' unstressed syllable durations differing significantly from those of Norwegians ($p = 0.041$).

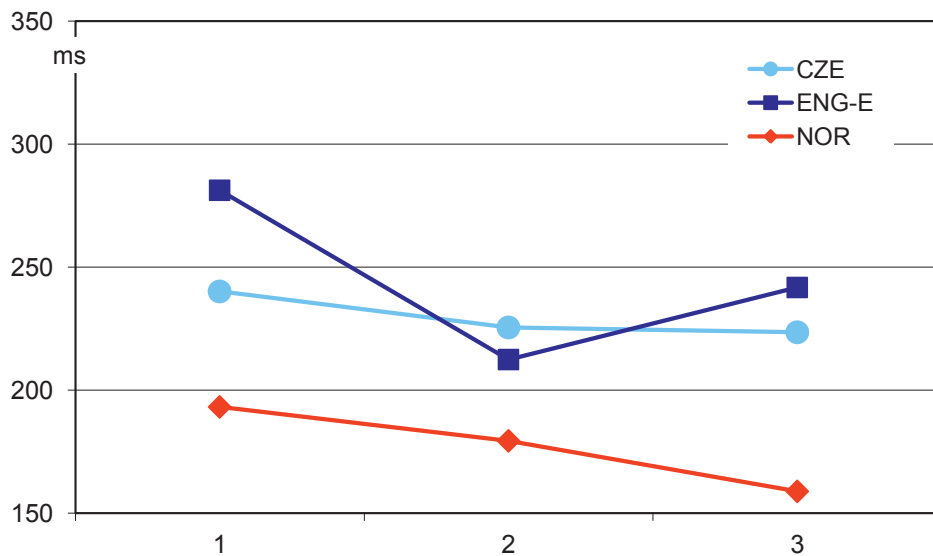


Figure 4.10: Mean unstressed syllable durations (in ms) of the first mention and two repeated mentions of polysyllabic content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E).

In the subset of *English data* (Figure 4.11), the mean unstressed syllable durations showed more noticeable shortening in repeated mentions, as compared to the first mentions of words (pooled across the items produced by speakers with different L1s it was 273 ms in first mentions vs. 206 and 221 in the repeated mentions of words). However, the shortening tendency was apparent especially in items spoken by native English speakers and Czech speakers. Items produced by Norwegian speakers also showed generally lower unstressed syllable durations (191 ms vs. 260 ms and 245 ms for items produced by Czech and native English speakers, respectively).

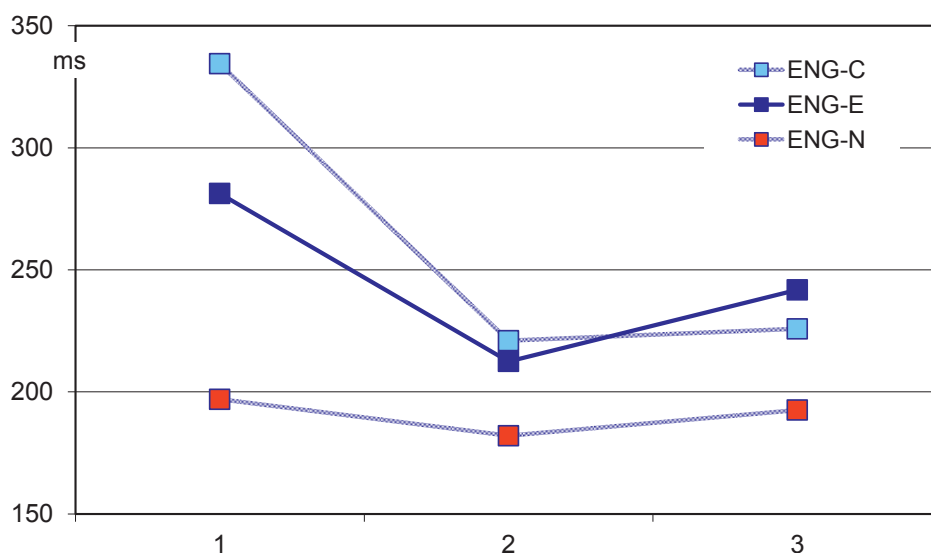


Figure 4.11: Mean unstressed syllable durations (in ms) of the first mention and two repeated mentions of polysyllabic content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).

A repeated-measures ANOVA with repeated mention as the within-group factor and speakers' native language as the between-group factor showed a significant effect of repeated mention ($F(2, 72) = 12.8$; $p < 0.001$) as well as an interaction between repeated mention and speakers' native language ($F(4, 72) = 3.14$; $p = 0.019$). The effect of speakers' L1 was not significant ($F(2, 36) = 1.70$; $p = 0.197$). The tests of within-subjects contrasts showed that the unstressed syllable duration in both the second and third successive mentions of words differed significantly from the first mention ($F(1, 36) = 18.4$; $p < 0.001$ and $F(1, 36) = 17.3$; $p < 0.001$, respectively). We may assume that the interaction is due to the occurrence of fairly long unstressed syllables in the first mentions of words and their consequent substantial shortening in repeated mentions in items produced by Czech speakers (and to a lesser degree also by native English speakers), contrasting with the nearly constant durations of unstressed syllables across the three mentions of words in productions of Norwegian speakers.

The last measure that we used to gain insight in the rhythmical properties of word realisations was the unstressed-to-stressed syllable duration ratio. Figure 4.12 shows the means of these ratios for the three successive mentions of content words comparing the three languages spoken by their native speakers, and Figure 4.13 shows the ratios in the three mentions of English words spoken by Czech, Norwegian and native English speakers. Quite surprisingly, in both figures we can observe that the ratios are mostly higher than 1, indicating that overall in the investigated (polysyllabic) content words, unstressed syllables had longer durations than the syllable with primary stress. Only the items in Norwegian and in English produced by Norwegian speakers seem to have on average approximately equal durations of stressed and unstressed syllables. When speculating about the causes of this unexpected result we should remember that the samples of polysyllabic items were rather small, especially for English spoken by all three speaker groups (see Table 4.3), and thus more susceptible to a bias due to the syllable structures of the lexical items represented in the sample. A closer inspection of the data showed that in the part of the data spoken by native English speakers, two lexical items, which had a syllable with secondary stress containing a tense vowel and a relatively complex structure (*beachballs*, *L-shape*), considerably influence the overall value of unstressed to stressed syllable duration ratio (mean ratio value for the 6 tokens was 2.2). In addition, two more words with a relatively complex unstressed syllable were identified (*letterbox*, *quarters*). In some of the realisations of these words (especially in pre-pausal position), the final fricative /s/ (or devoiced /z/) was considerably lengthened, influencing the resulting unstressed to stressed syllable duration ratio. The mean ratio value for the 12 tokens representing these 2 words was 1.5. The mean value of unstressed to stressed syllable duration ratio of the native English tokens after excluding the above described tokens was 0.8. In the subset of the data spoken by Norwegian speakers, one lexical item (*bucket*) was realised with a strongly affricated final /t/. The mean unstressed-to-stressed syllable duration ratio of three tokens of this word was 2.9, while the mean of the remaining items was 0.8. Similarly, among items spoken by Czech speakers, there were 3 words (*animals*, *elephant*, *woman*) where a simple and short stressed syllable was followed by a more complex unstressed syllable/s. In combination with a conspicuous lengthening of the

Chapter 4

final syllable (particularly in pre-pausal position), which seems to be typical for some of the Czech speakers, the mean unstressed-to-stressed syllable duration ratio of the 9 tokens representing these words reached a value of 2.8, while the mean ratio in the remaining items was 1.1. For a better overview, a complete list of the different lexical items for each language sorted by number of syllables can be found in Appendix D.

Figure 4.12 shows that whereas there are differences between the unstressed-to-stressed ratios in different languages (1.40, 1.19 and 1.00 for items in Czech, English and Norwegian, respectively), the ratios seem to be stable across the repeated mentions of words (1.16, 1.20 and 1.19 in the three mentions of words, respectively).

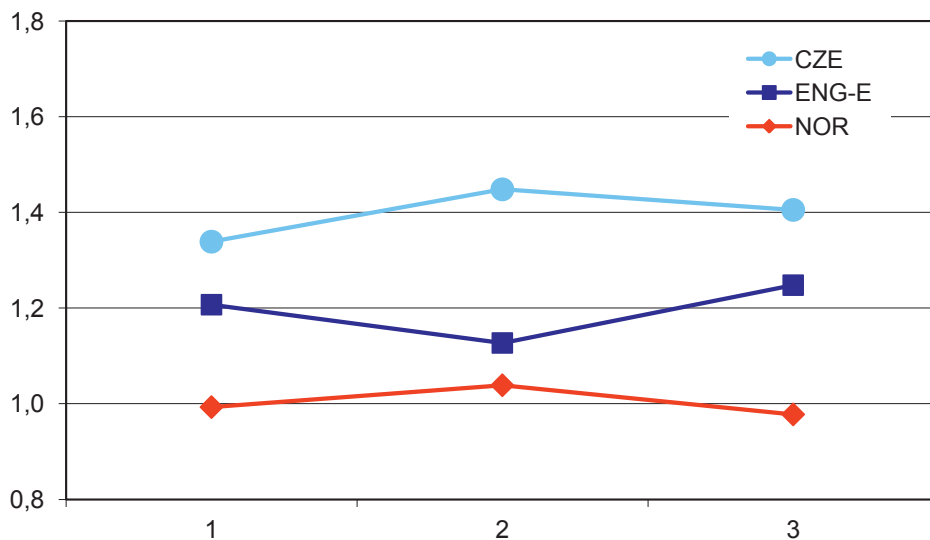


Figure 4.12: Unstressed-to-stressed syllable duration ratios in the first mention and two repeated mentions of polysyllabic content words in Czech (CZE), Norwegian (NOR) and English spoken by natives (ENG-E).

In the subset of *L1 data*, a repeated-measures ANOVA with repeated mention as the within-group factor and spoken language as the between-group factor confirmed a significant effect of the spoken language ($F(2, 63) = 3.60$; $p = 0.033$, while there was neither a significant effect of repeated mention, nor an interaction ($F(2, 126) < 1$ and

$F(4, 126) < 1$, respectively). Tukey's post-hoc test showed that the spoken language effect on the unstressed to stressed syllable duration ratio is due to Czech ratios being significantly higher (longer durations of unstressed relative to stressed syllables) than those of Norwegians ($p = 0.025$).

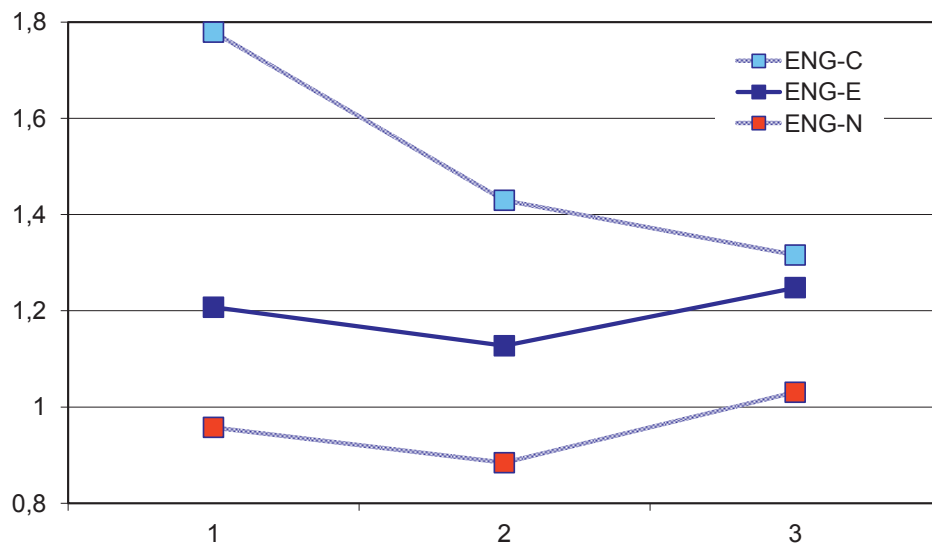


Figure 4.13: Unstressed-to-stressed syllable duration ratios in the first mention and two repeated mentions of polysyllabic content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).

In the subset of *English data* (Figure 4.13), apart from the differences between the unstressed-to-stressed ratios in different speaker groups based on speakers' L1 (1.51, 1.19 and 0.98 for English items spoken by Czech, English and Norwegian speakers, respectively), we can observe a differing pattern of unstressed-to-stressed ratio values in first and repeated mentions in the three groups based on speakers' L1. While in English items spoken by natives and Norwegian speakers the unstressed-to-stressed syllable duration ratio does not change much between the first and the two repeated mentions, Czech speakers have noticeably higher ratio values in first mentions of words. Surprisingly, a repeated-measures ANOVA with repeated mention as the within-group factor and speakers' L1 as the between-group factor showed, that neither of the main

Chapter 4

effects, nor their interaction reached significance ($F(2, 72) = 2.19$; $p = 0.119$, $F(2, 36) = 1.90$; $p = 0.164$ and $F(4, 72) = 2.24$; $p = 0.074$, respectively). However, when inspecting only the contrast between the first and third mention of the word, a significant interaction was found ($F(2, 36) = 4.61$; $p = 0.017$). As can be observed in Figure 4.13, this interaction seems to be due to Czech speakers' considerable drop in unstressed-to-stressed syllable duration ratio values between the first and third mentions of studied words as compared to the stable unstressed-to-stressed syllable duration ratios across all three mentions of words in items produced by native English speakers and Norwegians. This result is clearly related with the previously observed occurrence of very long unstressed syllables in Czech speakers' first mentions followed by a substantial shortening in repeated mentions, which was not paralleled by the patterns produced by speakers with other native languages. These findings, together with the lack of interaction in any other inspected durational variable (e.g. whole word durations, stressed syllable durations), indicate that in Czech speakers' English production, it was particularly the unstressed syllables of (polysyllabic) words that were prone to massive lengthening in first mentions, possibly due to final lengthening. Such a noticeable effect may be explained by an effort to hyperarticulate the first mentions of words, or may be an indication of uncertainty in production. Items produced by speakers with different L1s (i.e. native English speakers and Norwegians) did not follow this pattern, implying that any shortening (or lengthening) effect was spread evenly across stressed and unstressed syllables in the word.

4.3.1.4 Spectral reduction of repeated mentions

Unlike in the previous sections, where we inspected several variables relating to durational and rhythmical aspects of reduction, in this section we only used one variable to describe the spectral properties of the observed words. It was the measure of distance to the centroid calculated from the first two formants in the stressed vowel of the observed word (cf. Section 4.2.3.2; esp. Table 4.2 on page 170). A part of the triplets could not be used for this calculation (e.g. words where the stressed syllable contained a diphthong, syllabic /r/ or a strongly nasalised vowel). Apparently, the samples based on

spoken language and speakers' native languages also differed in terms of phonological identities of the represented vowels (i.e. the different words in the samples contained different vowels, although the vowels naturally remained constant across the three mentions of the words). In order to roughly compare the different phonetic qualities of vowels included in the samples, we assigned the occurring vowels to crude classes based on their phonological identity. The total numbers of triplets selected for the analysis of distance to the centroid, as well as the numbers of triplets containing vowels from each of these crude classes are listed in Table 4.4. Figure 4.14 shows the mean distance to the centroid in the stressed vowels of the three mentions of content words comparing items spoken by Czech, English and Norwegian native speakers. Figure 4.15 then shows the mean distance to the centroid in the stressed vowels of the three mentions of English words spoken by Czech, Norwegian and native English speakers. In both of the figures we can observe a slight but consistent decrease of the values in the repeated mentions of words, indicating that vowel qualities become more central in repeated mentions, compared to the words' first mentions.

The mean values across the *L1 data* subset were 2.6 Bark for first mentions and 2.5 and 2.4 Bark for the two repeated mentions, respectively. In the *English data* subset, the mean value of the distance to the centroid was 2.7 Bark for first mentions and 2.6 and 2.5 Bark for the two repeated mentions, respectively. Moreover, in Figure 4.14 we can observe considerably lower values of distance to the centroid in Czech production compared to Norwegian and native English productions (pooled across the three mentions, the mean distance to the centroid was 2.1 Bark for the Czech group vs. 2.7 and 2.5 Bark for English and Norwegian speakers, respectively). These differences may be partly due to uneven distributions of the vowels with different phonological identities across the samples for the individual groups (see Table 4.4). In addition, the properties of individual languages' vocalic systems should be taken into consideration. For example, the distance to the centroid of British English realisations of the tense vowel /ɔ:/ may be expected to be considerably greater compared to more fronted realisations of a comparable Czech phoneme /o:/.

Chapter 4

Table 4.4: Numbers of triplets containing vowels from the specified crude classes based on phonological identity, and total number of triplets used for analysis of distance to the centroid, in each group based on spoken language and speakers' L1.

	/i/	/ɪ/, /y/, /ʏ/	/e/, /ɛ/	/æ/	/a/, /ɑ/, /ʌ/	/o/, /ɔ/, /ɒ/	/u/, /ʊ/	Total
CZE	2	3	1	0	5	6	0	17
ENG-C	0	5	0	0	1	9	3	18
ENG-E	3	4	5	2	4	8	2	28
ENG-N	1	2	2	2	1	6	4	18
NOR	1	6	4	3	5	1	1	21

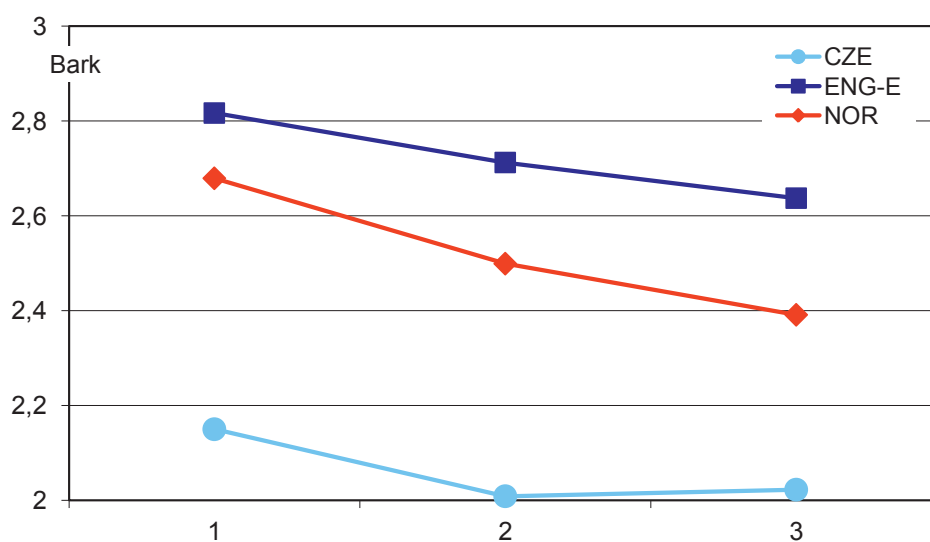


Figure 4.14: Mean distance to the centroid (in Bark) for stressed vowels in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E).

The mean values of distance to the centroid for vowels from each of the specified crude classes based on phonological identity, produced in the different groups based on spoken language and speakers' L1 are summarised in Figure 4.16. The displayed values should be taken with caution, since some of the means are calculated from a very small number of observations (see numbers of triplets in Table 4.4). Figure 4.16, for instance, clearly illustrates the difference in spectral contrast between the realisations of back

rounded vowels in different groups based on spoken language and speakers' L1. In our data, Norwegian items had the highest mean distance to the centroid in the back rounded vowels. The three groups of English items produced by speakers with different native languages had lower values of mean distance to the centroid, but the lowest mean distance to the centroid was observed in the group of Czech items (the mean was more than 2 Bark lower than the Norwegians' mean value). The resulting values for other vowel types show less contrast between the groups based on language and speakers' L1 (especially if we disregard cells with the lowest numbers of observations).

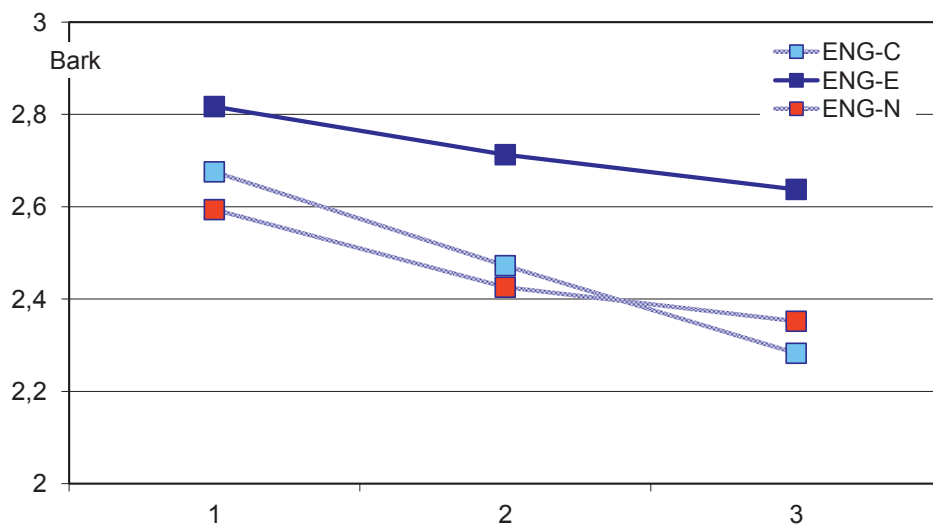


Figure 4.15: Mean distance to the centroid (in Bark) for stressed vowels in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).

With regard to the *L1 data* subset (Figure 4.14), a repeated-measures ANOVA with repeated mention as the within-group factor and spoken language as the between-group factor showed a significant effect of repeated mention ($F(2, 126) = 4.23$; $p = 0.017$) and a significant effect of spoken language ($F(2, 63) = 3.44$; $p = 0.038$), while the interaction of the main factors did not reach significance ($F(4, 126) < 1$). The tests of within-subjects contrasts showed that while the difference in the distance to the centroid between the

Chapter 4

words' first and second mentions did not reach significance ($F(1, 63)= 3.53$; $p= 0.065$), the difference between first and third mentions was significant ($F(1, 63)= 7.17$; $p= 0.009$). Tukey's post-hoc test showed that the spoken language effect on the distance to the centroid is due to Czech values being significantly lower than native English speakers' values ($p= 0.030$).

In the *English data* subset (Figure 4.15), a repeated-measures ANOVA with repeated mention as the within-group factor and speakers' native language as the between-group factor showed a significant effect of repeated mention ($F(2, 122)= 6.89$; $p= 0.001$) but neither the speakers' L1 nor its interaction with the repeated mention reached significance ($F(2, 61)< 1$ and $F(4, 122)< 1$, respectively). The tests of within-subjects contrasts showed that the distance to the centroid in both the second and third successive mentions of words differed significantly from the first mention ($F(1, 61)= 4.24$; $p= 0.044$ and $F(1, 61)= 11.5$; $p= 0.001$, respectively).

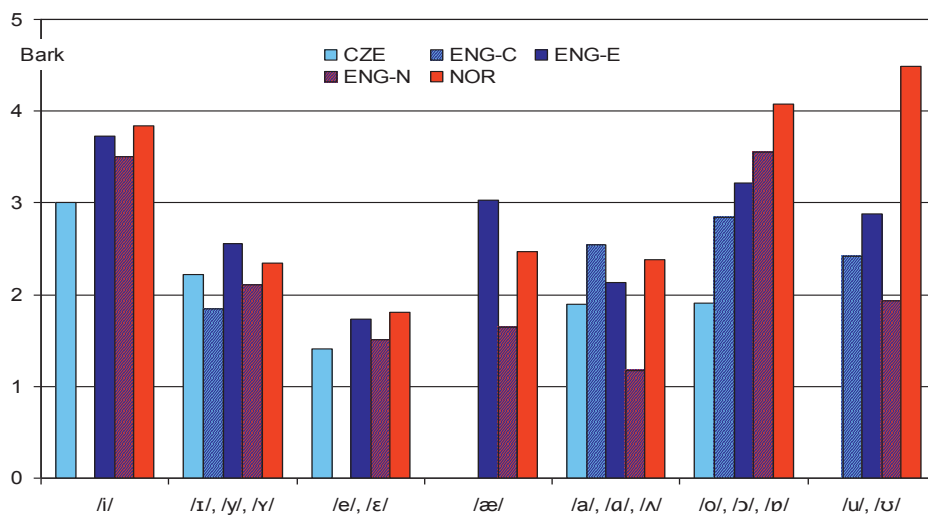


Figure 4.16: Mean distance to the centroid (in Bark) for vowels from each of the specified crude classes based on phonological identity, for the samples in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N).

4.3.2 Discourse-related control factor *Dialogue firstness*

This section describes the results of the analyses of effects of a discourse-related factor *dialogue firstness* on some of the measures of reduction in content words. This binary factor specifies whether the first member of a triplet was the very first occurrence of the given lexical item in the whole dialogue (see Section 4.2.3.3 for details). Due to the additional investigated factor in the analyses and the fact that the inspected samples are not particularly large, we restricted the analyses to the three variables that do not require splitting the data into further subgroups (i.e. item duration, mean syllable duration and vowel distance to the centroid). Still, the latter variable applies to a smaller subset of the data (i.e. only items containing monophthongs in their stressed syllables, excluding some problematic cases; see Section 4.2.3.2 for details).

The numbers of triplets whose first member was the first occurrence of the given word in the dialogue (these will be referred to as triplets with *dialogue-first* status) and their percentages within the total number of triplets in each group based on spoken language and speakers' L1 are listed in Table 4.5. We can observe that the percentages do not differ too much between the different groups, varying between 60% (English spoken by native speakers) and 75% (English spoken by Norwegians). It is also obvious that the numbers of triplets with non-first status are rather low (e.g. both the group of items in English spoken by Czech speakers and in English spoken by Norwegians only include 9 triplets with non-first status).

Table 4.5: Total numbers of triplets and numbers and percentages of triplets with dialogue-first status in each group based on spoken language and speakers' L1.

Language	Speakers' L1	Number of triplets	Dialogue-first status (triplets)	Dialogue-first status (%)
Czech	Czech	31	21	68
Norwegian	Norwegian	33	23	70
English	English	35	21	60
English	Czech	31	22	71
English	Norwegian	36	27	75

Chapter 4

Firstly, we will investigate the effect of dialogue firstness status on the whole word durations. Figure 4.17 shows the mean durations of the three mentions of content words comparing items in Czech, English and Norwegian spoken by native speakers, each group being split into subgroups depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue. Similarly, Figure 4.18 shows the mean word durations in the first mention and two repeated mentions of English words spoken by Czech, Norwegian and native English speakers, depending on their firstness status in the dialogue.

In Figure 4.17, we can observe that in the subgroups consisting of the triplets whose first member was the first occurrence of the given word in the dialogue (represented with closed markers and full lines), second and third mentions of words show a considerable decrease in duration as compared to first mentions (pooled across the groups based on spoken language, 509 ms for the first mentions vs. 416 ms and 401 ms for the two repeated mentions, respectively). To the contrary, the subgroups of triplets with non-first status (represented with open markers and dashed lines) do not indicate any shortening of the repeated mentions, compared to the triplets' first mentions (pooled across the groups based on spoken language, 397 ms, 415 ms and 406 ms for the three mentions, respectively).

In the *L1 data* subset, a repeated-measures ANOVA with repeated mention as the within-group factor and spoken language and dialogue firstness as the between-group factors showed a significant effect of repeated mention ($F(2, 186) = 5.84$; $p = 0.003$) and a significant interaction of repeated mention and firstness status ($F(2, 186) = 9.85$; $p < 0.001$), while the effects of spoken language and firstness status and other interactions did not reach significance.

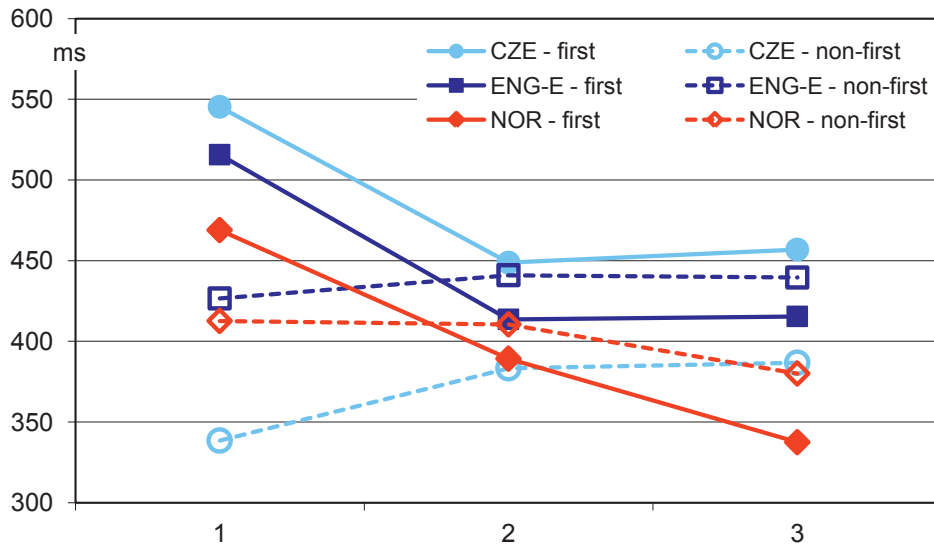


Figure 4.17: Mean durations (in ms) in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).

Figure 4.18 shows the situation in the *English data* subset, again indicating a tendency to durational decrease in repeated mentions in the subgroups formed by triplets with dialogue-first status (pooled across the groups based on speakers' native language, 535 ms for the first mentions vs. 445 ms and 428 ms for the two repeated mentions, respectively) as opposed to rather stable item durations in the three mentions in the subgroups formed by triplets with non-first status (pooled across the groups based on speakers' native language, 416 ms for the first mentions vs. 402 ms and 416 ms for the two repeated mentions, respectively). We may also observe that the tendency to stable item durations of the three mentions in non-first triplets is not quite followed by the group of items spoken by Czech speakers (with mean duration 470 ms for the first mention vs. 385 ms and 401 ms for the two repeated mentions, respectively).

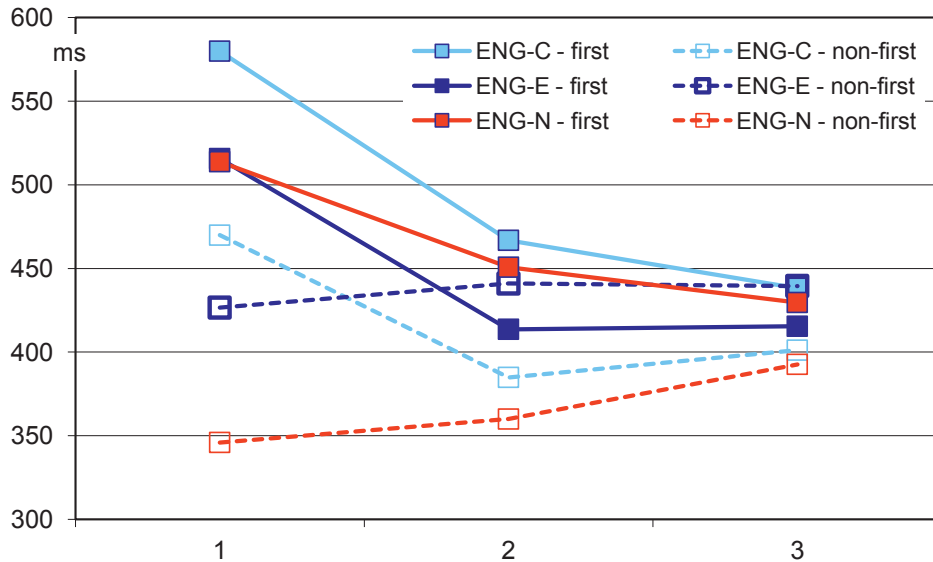


Figure 4.18: Mean durations (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).

A repeated-measures ANOVA with repeated mention as the within-group factor and speakers' native language and dialogue firstness status as the between-group factors showed significant effects of repeated mention ($F(2, 192) = 9.22$; $p < 0.001$), firstness status ($F(2, 96) = 6.32$; $p = 0.014$) and a significant interaction of repeated mention and firstness status ($F(2, 192) = 6.52$; $p = 0.002$). The effect of speakers' L1 and the other interactions did not reach significance. We may assume that the overall effect of firstness status, as well as the interaction with repeated mention, are largely due to the considerable difference in durations of first triplet members, depending on whether it was the first occurrence of the given word in the dialogue or not. Tokens selected as second and third mentions of the words by the speaker show much less noticeable difference as a result of the triplet's firstness status. In addition, to describe the interaction of dialogue firstness and repeated mention in more detail, we carried out a separate repeated-measures ANOVA for each of the three observed speaker groups based on speakers' L1 background. These analyses confirmed that while in the English

productions by natives and Norwegian speakers there is a significant interaction of dialogue firstness and repeated mention ($F(2, 66) = 3.49$; $p = 0.036$ and $F(2, 68) = 5.34$; $p = 0.007$ for the two speaker groups, respectively), in the English productions by Czech speakers, item duration is only affected by repeated mention ($F(2, 58) = 6.74$; $p = 0.002$) without a significant interaction with the firstness status ($F(2, 58) < 1$). This interesting finding can be interpreted as the Czech speakers' tendency to a more mechanical durational reduction of words, once they are produced again by the same speaker, as opposed to English and Norwegian speakers applying reduction selectively, depending on the overall word's status in the discourse.

The next measure used to inspect the effect of dialogue firstness status on the reduction of repeated mentions of content words was mean syllable duration. Figure 4.19 shows the mean syllable durations of the three mentions of content words comparing items in Czech, English and Norwegian spoken by native speakers, each group being split into subgroups depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue. Similarly, Figure 4.20 shows the mean syllable durations in repeated mentions of English words spoken by Czech, Norwegian and native English speakers, depending on their firstness status in the dialogue. The values of syllable duration follow a similar pattern as whole word durations: both in the subset of *L1 data* and *English data* the items with first status show a decrease of syllable duration in second and third mention, while the items with non-first status do not display a decrease in syllable duration. The exception is, again, the group of items in non-native English spoken by Czech speakers, where the mean syllable duration decreases in repeated mentions in all items, regardless of the firstness status (see Figure 4.20). Moreover, in Figure 4.19 we can observe a large effect of spoken language, which could be expected based on previous results for the effects of repeated mention and spoken language (cf. Figure 4.4).

In the *L1 data* subset (Figure 4.19), a repeated-measures ANOVA with repeated mention as the within-group factor and spoken language and dialogue firstness status as the between-group factors showed a significant effect of repeated mention ($F(2, 186) =$

Chapter 4

3.40; $p= 0.035$) and a significant interaction of repeated mention and firstness status ($F(2, 186)= 9.52$; $p< 0.001$) while the interaction of repeated mention and spoken language did not reach significance. Since Levene's test indicated unequal variances with high significance ($p< 0.001$) we performed the statistical test of the between-subject effect of firstness status using a non-parametric test. The Kruskal-Wallis test showed no significant effect of firstness status on mean syllable duration ($\chi^2(1)= 1.59$; $p= 0.207$). Clearly, as we tested in Section 4.3.1.2, there was also a significant effect of spoken language.

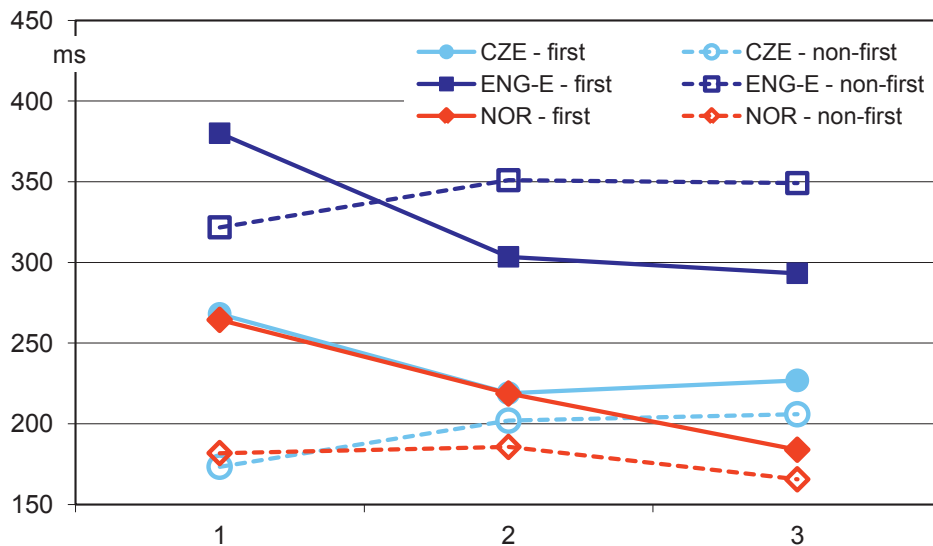


Figure 4.19: Mean syllable durations (in ms) in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).

In the *English data* subset (Figure 4.20), a repeated-measures ANOVA with repeated mention as the within-group factor and speakers' native language and firstness status as the between-group factors showed a significant effect of repeated mention ($F(2, 192)= 5.64$; $p= 0.004$) and a significant interaction of repeated mention and firstness status

($F(2, 192) = 6.71$; $p = 0.002$). The effects of speakers' L1 background, firstness status and the other interactions did not reach significance.

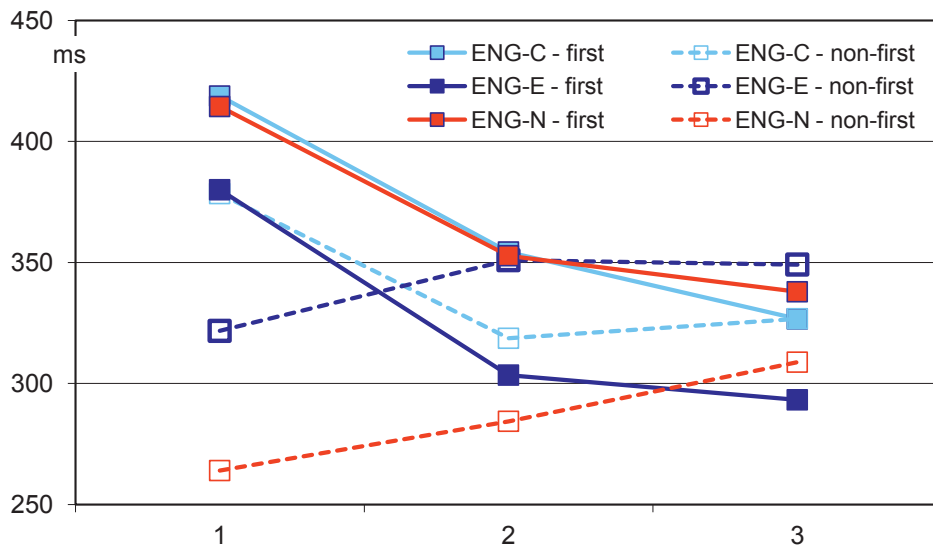


Figure 4.20: Mean syllable durations (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).

Similarly as in the analysis of item duration, we attempted to describe the interaction of dialogue firstness and repeated mention reduction in more detail by carrying out separate repeated-measures ANOVAs for the three observed groups based on speakers' L1 background. As in the case of word duration, here the analyses also confirmed differences between the patterns produced by Czech speakers and the other two speaker groups. In the English words produced by Czech speakers, syllable duration decreased in repeated mentions regardless of firstness status ($F(2, 58) = 4.37$; $p = 0.017$). Both in the native English speaker group and in the Norwegian speaker group, there was no main effect of repetition on mean syllable duration, but a significant interaction of

Chapter 4

dialogue firstness and repeated mention was found ($F(2, 66) = 4.09$; $p = 0.021$ and $F(1.84, 62.546) = 6.10$; $p = 0.005$, for English and Norwegian⁴⁷ group, respectively).

Lastly, we inspected the measure of vowel distance to the centroid in the subset of the data containing suitable monophthongs in their stressed syllable (for details see Section 4.2.3.2). Figure 4.21 shows the mean distance to the centroid in the stressed vowels of the three mentions of content words comparing items in Czech, English and Norwegian spoken by native speakers, each group being split into subgroups depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue. Similarly, Figure 4.22 shows the mean distance to the centroid in the stressed vowels of the three mentions of English words spoken by Czech, Norwegian and native English speakers, depending on their firstness status in the dialogue.

In these figures, the subgroups of items with dialogue-first status show a very slight decrease of mean distance to the centroid in the repeated mentions, as compared to first mentions, while no difference between first and repeated mentions in the non-first items can be observed. In the *L1 data* subset (Figure 4.21), a repeated-measures ANOVA with repeated mention as the within-group factor and spoken language and firstness status as the between-group factors showed only a significant effect of spoken language ($F(2, 60) = 3.71$; $p = 0.030$). In addition, there was a marginally significant effect of repeated mention ($F(2, 120) = 2.97$; $p = 0.055$). The effect of dialogue firstness status and any interactions did not reach significance.

⁴⁷ Mauchly's test of sphericity indicated that the sphericity assumption had been violated, $\chi(2) = 6.25$; $p = 0.044$, and therefore, a Huynh-Feldt correction was used.

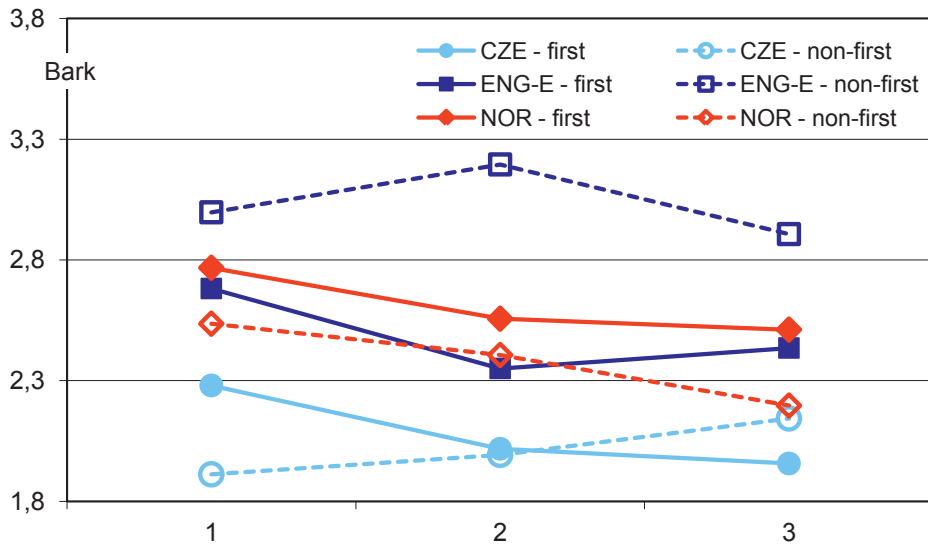


Figure 4.21: Mean distance to the centroid (in Bark) for stressed vowels in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).

In the *English data* subset (Figure 4.22), a repeated-measures ANOVA with repeated mention as the within-group factor and speakers' native language and firstness status as the between-group factors showed only a significant interaction of repeated mention and firstness status ($F(2, 116) = 3.63$; $p = 0.030$). The main effects of repeated mention, speakers' L1 and firstness status as well as the other interactions did not reach significance.

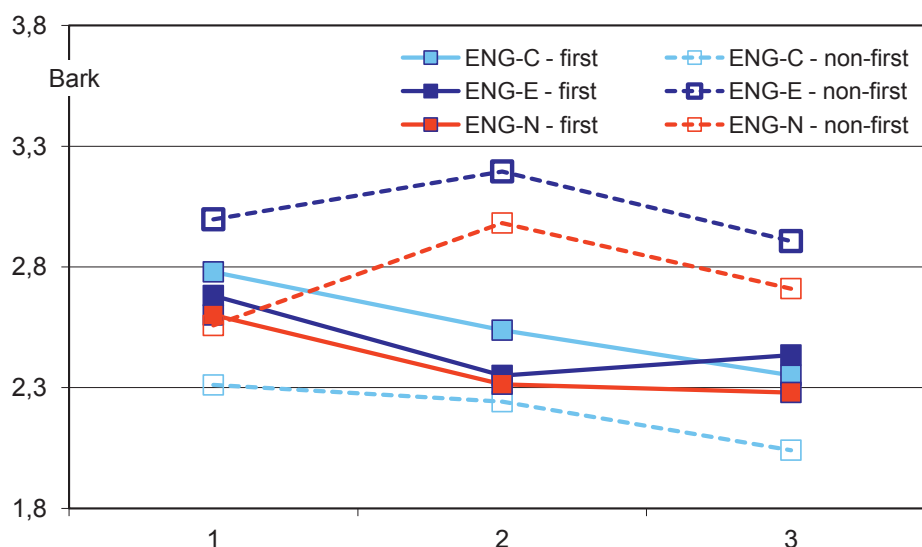


Figure 4.22: Mean distance to the centroid (in Bark) for stressed vowels in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on whether the first member of the triplet was the first occurrence of the given word in the dialogue (first) or not (non-first).

4.3.3 Control factors related to prosodic structure

As explained in Section 4.2.3.3, two factors related to prosodic structure were controlled for: the presence of prosodic prominence in the observed word, and the word's position in relation to major prosodic boundaries. The presence of prosodic prominence (accentuation) was evaluated by three listeners (for details, see Section 4.2.3.3). The results of the prominence evaluation are presented in Table 4.6 which lists the total numbers of observed words and numbers and percentages of tokens evaluated as accented for each group of items based on spoken language and speakers' L1.

Since the results of the evaluation showed that almost all the observed tokens carry a certain degree of prosodic prominence, while only a negligible number of tokens were judged as deaccented, we can assume that our data are rather homogeneous from this point of view. Therefore, it is not necessary to carry out any further analyses to establish the influence of this factor on the studied variables.

Table 4.6: Total numbers of observed word tokens and numbers and percentages of tokens evaluated as accented in each group based on spoken language and speakers' L1.

Language	Speakers' L1	Number of tokens	Accented (tokens)	Accent (%)
Czech	Czech	93	93	100
Norwegian	Norwegian	99	99	100
English	English	105	102	97
English	Czech	93	93	100
English	Norwegian	108	106	98

The other control factor related to the prosodic structure examined in this study is the binary variable *finality*, coding the presence of a major prosodic boundary following the observed word. As described in Section 4.2.3.3, tokens were coded as *final* when followed by a silent pause or a strong prosodic disjuncture. The distribution of word occurrences in final and non-final position for the first and two following mentions of words in the groups based on spoken language and speakers' L1 are displayed in Figure 4.23.

We can observe that in almost all the groups, first mentions of content words occur more often in final positions than repeated mentions. Although we can observe somewhat higher proportions of final tokens in English, and particularly in the two groups spoken by non-native speakers, when considering the two subsets of data separately, the differences between the distributions of final items in the different speaker groups are not striking. Pooling across the words' mentions, Pearson's Chi-squared tests showed that both within the *L1 data* subset and within the *English data* subset there were no significant differences between the finality distribution in the speaker groups ($\chi^2(2) = 4.16$; $p = 0.125$ and $\chi^2(2) = 0.810$; $p = 0.667$, for *L1 data* and *English data*, respectively). The somewhat higher proportions of final tokens in English spoken by Czech speakers and English spoken by Norwegians may result from the non-native speakers' lower fluency (e.g. increased number of pauses).

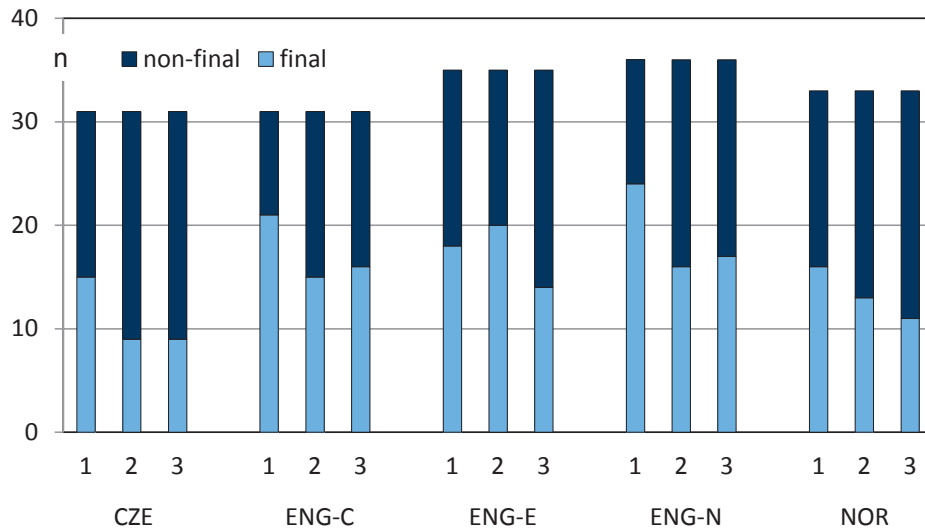


Figure 4.23: Distributions of occurrence of words in final and non-final position within a higher prosodic unit (numbers of tokens) for the first mention and two repeated mentions of words in the groups based on spoken language and speakers' L1.

Due to the fact that the value of the control variable *finality* varies not only between but also within the triplets, we were not able to use a repeated-measures ANOVA to test its effect on various measures of reduction. Instead, we used mixed-effect linear models, which seem to offer a well-suited method for determining the effect of this factor on the observed measures of reduction. Mixed-effect models describe the variation in the data induced by the random effects (for example speaker or item identity) in parallel with modelling the contributions of the fixed factors. Since this method allows us to disregard the variation due to the different words included in the samples, there is no need to use additional measures of duration apart from the whole word duration (cf. Section 4.3.1.2). Therefore, in this section we will only address whole word duration and vowel distance to the centroid.

For each of the observed dependent variables, we built a model in the following way. After determining which of the random effects (word, speaker) contribute significantly to the model goodness of fit for a given dependent variable, fixed effects are added

using a stepwise inclusion procedure. The fixed factors tested included the spoken language (for the *L1 data* subset) / speakers' L1 (for the *English data* subset), repeated mention, the logarithmically transformed articulation rate and the binary variable finality. In each step, the current "best" model is compared with models built from that model by including one more of the so far unused factors, or an interaction of the factors already included in the model. As long as at least one of the extended models has a significantly better fit than the current "best" model, the model achieving the lowest values of information criteria AIC and BIC is adopted as the next "best" model. Detailed results of statistical analyses carried out to find the best-fitting model for the dependent variables item duration (Dur.item) and distance to the centroid (Centr.dist) in the *L1 data* subset can be found in Appendix E. Appendix F then contains the details of statistical analyses run in order to fit the models for these dependent variables in the *English data* subset.

As for item duration, in the *L1 data* subset the best model was found to contain only word identity as a random factor and repeated mention, finality, logarithmically transformed articulation rate and the interaction between finality and logarithmically transformed articulation rate as fixed factors. Figure 4.24 uses residuals after modelling the random effect of word identity to plot the effect of repeated mention and finality on durations of content words in the groups based on spoken language. We can observe that items occurring in final positions in a prosodic unit have generally longer durations. The difference is somewhat smaller for items in Norwegian. At the same time, a consistent shortening of repeated mentions can be observed both for final and non-final items. The effect of the logarithmically transformed articulation rate and its interaction with finality was closer inspected by examining correlations between articulation rate and duration residuals. It turned out that the correlation in the whole *L1 data* set was highly significant but only achieved a weak negative correlation coefficient ($r = -0.211$; $N = 295$; $p < 0.001$), meaning that a mere 4.5% of the variation in the durational data can be explained by articulation rate. When we inspected the data split into two groups depending on the finality status, we found that the correlation of durational residuals with articulation rate in the subset of final items is not significant ($r = -0.063$; $r^2 = 0.004$;

Chapter 4

$N= 123$; $p= 0.487$), while in the non-final items subset, there is a moderate negative correlation which is highly significant ($r= - 0.308$; $r^2= 0.095$; $N= 172$; $p< 0.001$). In other words, in non-final position, higher articulation rates in immediate surroundings of a word are associated with shorter words durations. In final positions, on the other hand, the relationship of word duration and articulation rate is not present.

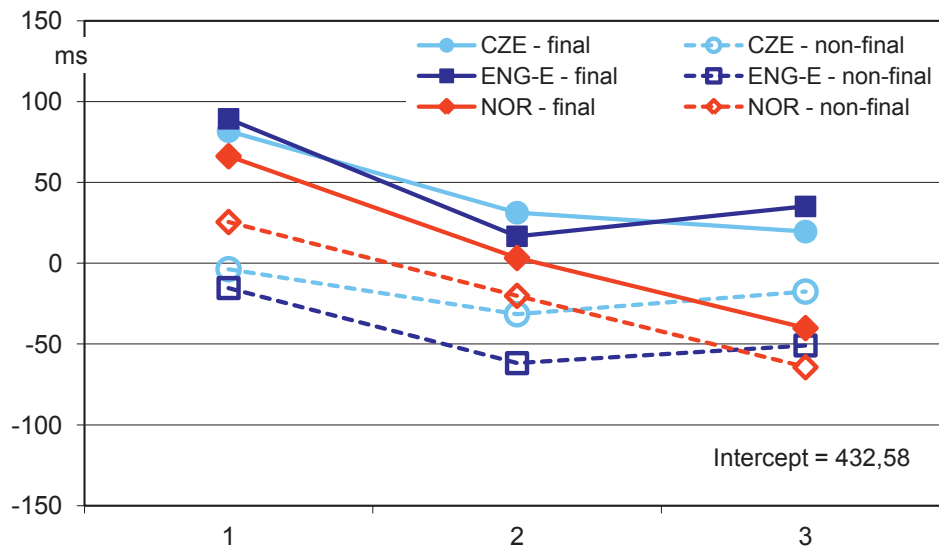


Figure 4.24: Word duration residuals after modelling the random effect of word identity (in ms) in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on the word's position as final or non-final in a prosodic unit.

In the *English data* subset, both word and speaker identity contributed significantly to the model fit as random factors, while the significant fixed factors included repeated mention, finality and logarithmically transformed articulation rate. Figure 4.25 uses residuals after modelling the random effect of word and speaker identity to plot the effect of repeated mention and finality on durations of content words in the groups based on speakers' native language. As in the *L1 data* subset, items in final positions have generally longer durations. This difference is less noticeable in items produced by Norwegian speakers. In addition, there is an observable effect of repeated mention on

all groups of items. The correlation between logarithmically transformed articulation rate and duration residuals was found to be a highly significant, but weak negative correlation ($r = -0.185$; $r^2 = 0.034$; $N = 302$; $p = 0.001$).

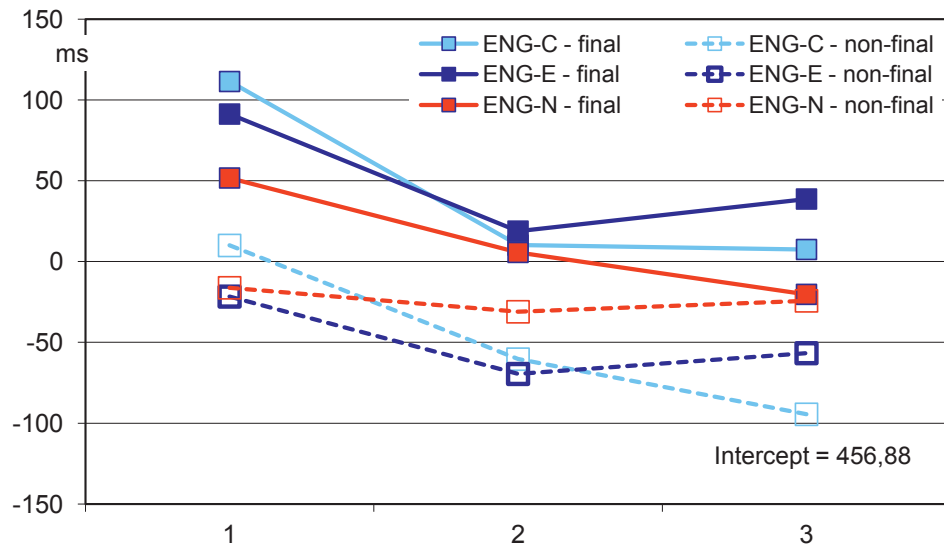


Figure 4.25: Word duration residuals after modelling the random effects of word and speaker identity (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on the word's position as final or non-final in a prosodic unit.

Figures 4.26 and 4.27 use residuals after modelling the random effect of word and speaker identity in the respective subsets of data to plot the effect of repeated mention and finality on the spectral contrast in the stressed vowels of the observed content words. In the model for the *L1 data* subset, apart from word and speaker identity as random factors, the significant fixed factors included repeated mention and logarithmically transformed articulation rate. Figure 4.26 shows the residuals of vowel distance to the centroid in the groups based on spoken language. We can observe that the mean values of vowel distance to the centroid decrease with repeated mention, while this measure of spectral contrast does not seem to be noticeably affected by the position of the word as final or non-final in a prosodic unit.

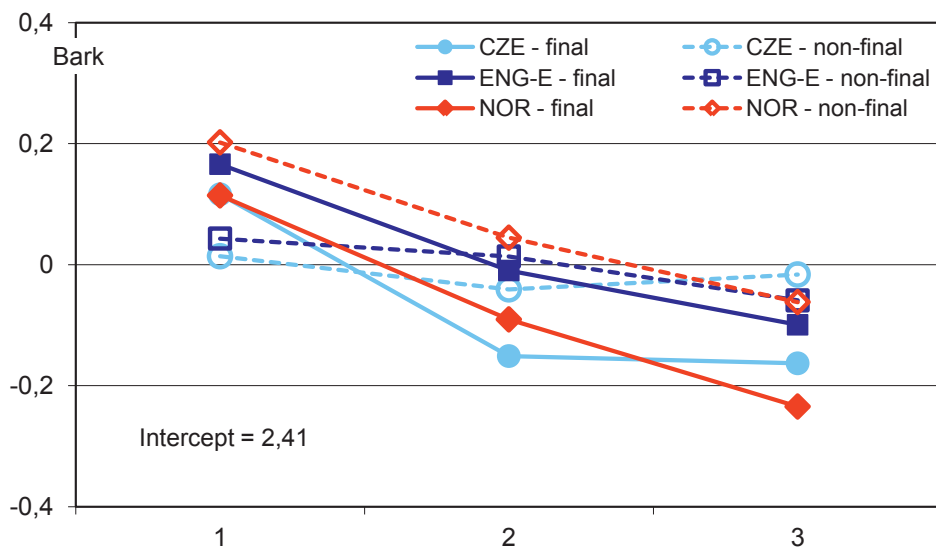


Figure 4.26: Distance-to-centroid residuals after modelling the random effects of word and speaker identity (in Bark) in the first mention and two repeated mentions of content words in Czech (CZE), Norwegian (NOR) and English spoken by native speakers (ENG-E), depending on the word's position as final or non-final in a prosodic unit.

In the *English data* subset, both word and speaker identity contributed significantly to the model fit as random factors, while the significant fixed factors included repeated mention and speakers' L1 background. Mean values of distance-to-centroid residuals in the *English data* subset are displayed in Figure 4.27. Similarly as in the *L1 data*, distance to the centroid slightly decreases with repeated mention, but overall it is not much influenced by the position of the word as final or non-final in a prosodic unit. There are also no overall differences between the spectral qualities of items produced by speakers with different L1 backgrounds. However, in the subset of items produced by Czech speakers we can observe a noticeable fall of the mean distance to the centroid in the third mentions in non-final positions. A closer inspection of the data revealed that these results are based on rather few (9) observations. In some of these items, we can observe a fair decrease of vowel spectral contrast (as compared to earlier mentions of the same word produced by the same speaker). However, a very conspicuous decrease of vowel distance to the centroid (as compared to earlier mentions of the word) was

observed in one of the items. Within the triplet, the difference of vowel distance to the centroid between the first and third mentions of the word amounted to 2.5 Bark. It seems that the resulting very low value in non-final third mentions produced by Czech speakers is due to a strong influence that this one observation had on the rather small group of items.

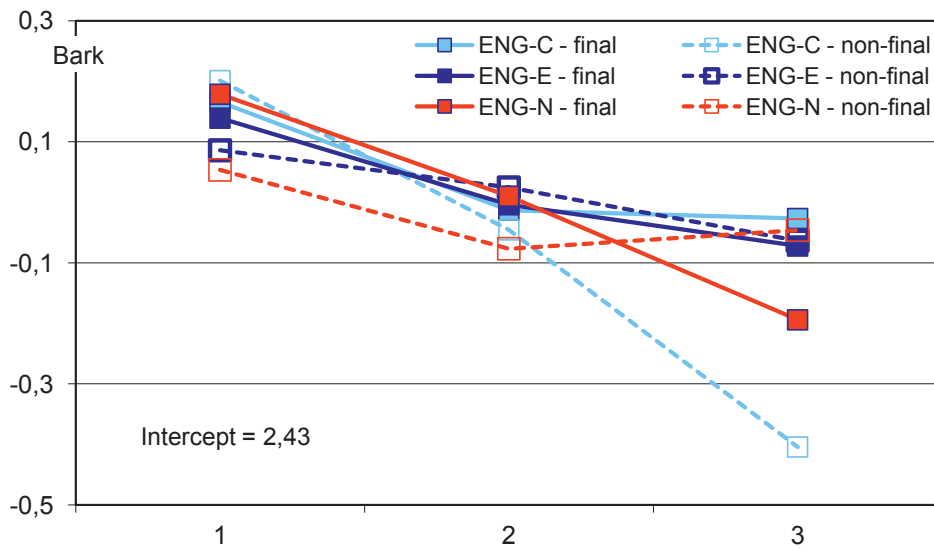


Figure 4.27: Distance-to-centroid residuals after modelling the random effects of word and speaker identity (in ms) in the first mention and two repeated mentions of content words in English spoken by native speakers (ENG-E), by Czech speakers (ENG-C) and by Norwegians (ENG-N), depending on the word's position as final or non-final in a prosodic unit.

4.4 Summary and discussion

In the following two subsections we will summarise and discuss the most important findings of the analyses of the main experimental factors repeated mention and speaker group (Section 4.4.1), and give an overview of the observed effects of control factors included in our study (Section 4.4.2).

4.4.1 Effect of repeated mention and speaker group

The following Table 4.7 summarises the results of all statistical analyses concerning the effects of repeated mention and speaker group (i.e. groups based on spoken language or speakers' L1 backgrounds) on various inspected measures of phonetic reduction. In the left column, the results pertaining to analyses on the *L1 data* subset are listed, while the right column summarises analyses on the *English data* subset. The cells corresponding to measures that were not relevant in a given subset of items are crossed out. In the following paragraphs, the statistically significant results will be discussed.

As can be seen, both in the subset of *L1 data* and in the subset of *English data*, the effect of repeated mention on whole word duration was highly significant, while neither an effect of spoken language or speakers' L1 background (for each subset respectively), nor an interaction with repeated mention reached significance. Similarly, repeated mention was found to have a significant effect on mean syllable duration in both subsets. In addition, in the *L1 data* subset, spoken language was found to have a significant effect on mean syllable duration. In particular, the mean syllable duration in English words was considerably greater than syllable durations in Czech and Norwegian words. This finding is not surprising and can be explained by the different proportion of monosyllabic and polysyllabic words in the samples in the three languages. While in the English samples more than half of the inspected words were monosyllabic (i.e. only contained the stressed syllable), in Czech and Norwegian, polysyllabic words with expectedly shorter mean syllable durations formed a majority of the sample (see Table 4.3 on page 181 for details). In sum, both of these durational variables show a considerable shortening in repeated mentions as compared to the first word's mention. This is consistent with the findings of a number of studies (e.g. Fowler and Housum, 1987; Shields and Balota, 1991; Baker and Bradlow, 2009). Some of the studies showed that this tendency is particularly noticeable in communicative contexts (Fowler, 1988; McAllister et al., 1994). The "picture replication task" used for the elicitation of the spontaneous dialogues used in this study seems to provide well-suited material for investigation of these phenomena. Although most of the previous studies were carried out on native English material, the shortening of repeated mentions was confirmed even

in other languages (Baker and Bradlow, 2007) and in non-native English (Baker et al., 2011). This is consistent with our results both in the *L1 data* and *English data* subsets.

Table 4.7: Results of all statistical analyses concerning the effects of repeated mention (repetition) and speaker group based on spoken language or speakers' L1 background.

	variable	L1 data subset		English data subset	
Durational measures	Whole word duration	Repetition ***		Repetition ***	
	Mean syllable duration	Repetition ** Spoken language ***		Repetition ***	
		monosyllabic words	polysyllabic words	monosyllabic words	polysyllabic words
Rhythmical aspects of reduction	Stressed syllable duration	Repetition . Spoken language *	Repetition ***	Repetition **	Repetition **
	Unstressed syllable duration	/	Repetition ** Spoken language *	/	Repetition *** Speakers' L1:repetition *
	Unstressed-to-stressed syllable duration ratio	/	Spoken language *	/	(only contrast of 1 st and 3 rd mention) (Speakers' L1:repetition *)
Spectral measures	Distance to the centroid	Repetition * Spoken language *		Repetition **	

Chapter 4

As for further duration-based measures which refer to words' rhythmical aspects, the effect of repeated mention on the duration of stressed syllables (analysed separately in the subsets of monosyllabic and polysyllabic words) and mean duration of unstressed syllables (obviously only in the subset of polysyllabic words) was for the most part consistent and significant in both the *L1 data* and *English data* subsets. Only in the subset of monosyllabic words in the *L1 data* subset, did the effect of repeated mention on stressed syllable duration achieve merely a marginal significance. This could, however, be due to the smaller size of this sample. In Czech and Norwegian, monosyllabic words only formed a minor part of the sample, amounting to as few as 7 monosyllabic triplets for Czech and 5 for Norwegian. Since the tendency was otherwise similar to those observed for other variables, we assume that with a larger sample, the effect would reach significance. These results further confirm the tendency to the durational reduction of repeated mentions, which is consistent with a number of studies mentioned above. More interestingly, a significant effect of spoken language on stressed syllable duration in monosyllabic L1 words was found. This was due to the fact that the durations of stressed syllables were significantly longer in English words than in Czech words. This finding may possibly be explained by the relatively different rhythmical properties of the two languages. While English as a conspicuously stress-timed language uses duration as one of the means to achieve prominence in stressed syllables, the vowel length contrast in Czech does not enable excessive use of duration for signalling stressed syllables (see Section 1.3.1 for details). Further, an effect of spoken language on unstressed syllable duration was found, consisting in the English unstressed syllables being significantly longer than the Norwegian ones. This may be explained by a stronger reduction of Norwegian unstressed syllables, but may also be partly due to syllable structures of the lexical items represented in the sample, in particular its English part (for more details about the composition of samples of polysyllabic words see pages 192-193 or Appendix D). As for the mean unstressed syllable duration in the subset of *English data*, apart from the main effect of repeated mention, a significant interaction of speakers' L1 background and repeated mention was found. This interaction is possibly due to the occurrence of fairly long unstressed syllables in the first mentions of words and their consequent substantial shortening in repeated mentions in items produced by Czech speakers (and to a lesser degree also by native English

speakers), contrasting with nearly constant unstressed syllables durations across the three mentions of words in productions of Norwegian speakers. This may be the result of a noticeable final lengthening or inappropriate hyperarticulation of unstressed syllables in the first mentions of words by Czech speakers, and may therefore be seen as a consequence of different rhythmical characteristics of the speakers' L1.

A different way to investigate the rhythmical aspects of repeated mentions reduction is using the unstressed-to-stressed syllable duration ratio. This measure describes the relation of stressed and unstressed syllables' durations and thus characterises the contributions of either to the overall durational reduction of the repeated mentions of words. Our results showed that this measure does not show any main effect of repeated mention in either of the studied subsets. In other words, any durational changes due to words' repeated mentions seem to affect the stressed and unstressed parts proportionally. However, in the subset of *L1 data*, a significant effect of spoken language on the ratios was found. In particular, Czech items were found to have significantly higher ratios compared to items in Norwegian, indicating overall relatively longer durations of unstressed syllables in Czech material. As for the *English data* subset, although the overall interaction between speakers' L1 background and repeated mentions did not reach significance, when considering only the contrast between the first and third mentions, Czech speakers showed a significantly different pattern than the other two speaker groups. While in English items spoken by the natives and Norwegian speakers the unstressed-to-stressed syllable duration ratio does not change between the first and third mention, Czech speakers have noticeably higher ratio values in first mentions of words. This suggests that the durational reduction of repeated mentions in English items spoken by Czech speakers is achieved largely by the massive shortening of unstressed syllables. Obviously, this result is related to the previously discussed occurrence of unusually long (hyperarticulated) unstressed syllables in the first mentions of English items produced by Czech speakers and is possibly related to different rhythmical characteristics of the speakers' L1 (Czech) as compared to English (cf. Section 1.3.1). Moreover, deviations from native-like rhythmical patterns, as expressed using a ratio of stressed and unstressed syllable durations, are in accordance

Chapter 4

with the results of a number of previous studies (e.g. Gut, 2007; Volín, 2005; Trofimovich and Baker, 2006). The above mentioned interaction then points out the first mentions as a condition where Czech speakers of English deviate particularly strongly from the native-like pattern. In combination with the fact that the native productions in Czech also seem to keep stable unstressed-to-stressed syllable duration ratios across all three mentions of words, the conspicuously lengthened unstressed syllables may be due to the hesitant manner of production⁴⁸ of the first mentions of English words by Czech speakers.

Lastly, the only measure describing the spectral properties of the observed words was the distance to the centroid calculated from the first two formants in the stressed vowel. Also this measure showed a significant effect of repeated mention (i.e. a decrease of spectral contrast in repeated mentions of words) in both the *LI data* and *English data* subsets. This result is consistent with previous research reporting spectral contrast reduction in repeated mentions in Dutch (Koopmans-van Beinum and van Bergem, 1989), but the consistent pattern of spectral contrast reduction in all observed languages, as well as in English produced by non-native speakers, contributes to the current knowledge by confirming the generality of spectral contrast reduction in the stressed vowels in repeated mentions of content words. Apart from the effect of repeated mention, in the *LI data* subset, we also found a significant effect of spoken language. This is not surprising considering the differences between the vowel systems and the expected qualities of vowels in the three observed languages.

4.4.2 Control factors

This section will summarise and discuss the findings relating to the effects of the control factors addressed in this study on some aspects of phonetic reduction. The following Table 4.8 lists the results of statistical analyses concerning the effect of

⁴⁸ In most of the cases, the observed word was in final position. Moreover, speakers seemed to take their time with the production of the given intonational phrase rather than rush into continuing their utterance.

control factor dialogue firstness and the main experimental factors repeated mention and speaker group on selected measures of phonetic reduction. In the left column the results pertaining to analyses on the *L1 data* subset are listed, while the right column summarises analyses on the *English data* subset. In the following paragraphs, the statistically significant results will be discussed.

Table 4.8: Results of statistical analyses concerning the effects of repeated mention (repetition), speaker group based on spoken language or speakers' L1 background and dialogue firstness.

variable	L1 data subset	English data subset
Whole word duration	Repetition ** Repetition:firstness ***	Repetition *** Firstness * Repetition:firstness ** (Czech speakers: no interaction)
Mean syllable duration	Repetition * Spoken language *** Repetition:firstness ***	Repetition ** Repetition:firstness ** (Czech speakers: no interaction)
Distance to the centroid	Repetition . Spoken language *	Repetition:firstness *

As we can observe in the table above, even after including the factor dialogue firstness in the analysis of variance, the effect of repeated mention remains significant in both of the durational measures for both investigated data subsets. Moreover, the mean syllable duration is significantly influenced by spoken language in the *L1 data* subset, just like in the analyses of main experimental factors. Apart from that, a significant interaction of repeated mention and firstness is found to affect all inspected variables in the *English data* subset, and both durational variables in the *L1 data* subset. This interaction confirms that the effect of repeated mention differs depending on whether the speaker's first mention of a word is also its first occurrence in the dialogue: while the triplets containing the word's first occurrence in the dialogue show consistent reduction

Chapter 4

(durational or spectral), the other⁴⁹ triplets follow a different pattern. This tendency is consistent with the results of a study by Bard et al. (2000), showing a shortening and intelligibility decrease of second mentions of words in Map Task dialogues regardless of which participant produced the first mention. A considerable shortening of repeated mentions of words relative to their first mentions by a given speaker can therefore be only expected in the triplets with dialogue-first status. However, an exception to this tendency was revealed in whole word and syllable durations in English items produced by Czech speakers. It turns out that the Czech speakers' English productions show durational reduction even in items where the first triplet member was not the word's first occurrence in the dialogue. This may indicate Czech speakers' limited ability to adjust the durational properties in their English productions in relation to a word's overall discourse status, resulting in a more mechanical durational reduction of words once they are produced repeatedly by the same speaker. Bard et al. (2000) suggest that duration and intelligibility of repeated mentions are simultaneously controlled by two types of mechanisms: fast automatic priming processes only dependent on speaker's knowledge and slower, more cognitively demanding processes that involve inference about the listener's knowledge. We may therefore speculate that this pattern of findings is due to Czech speakers' inability to efficiently deal with the cognitive demands posed by the conversational task (carried out in L2) that hinders their use of such listener-directed adaptive processes.

As for the vowel distance to the centroid in the *L1 data* subset, no significant interaction of repeated mention and firstness was found, and the effect of repeated mention itself was only marginally significant. As expected, a significant effect of spoken language was found as well.

As to the effects of control factors related to prosodic structure, we originally intended to inspect two factors: the prosodic prominence in the observed word, and a word's

⁴⁹ Triplets, where the speaker's first mention of a word was not the word's first occurrence in the dialogue; i.e. the word had been mentioned previously by the other speaker.

position in relation to major prosodic boundaries. An analysis of the distribution of prosodic prominence in the observed words as evaluated by three listeners showed, however, that in our sample nearly all the observed words are perceived as accented. Therefore, we did not consider it necessary to carry out any further analyses investigating the effect of prominence on the observed words.

The results of the statistical analyses carried out to establish the effects of the control factors finality and articulation rate, as well as the main experimental factors repeated mention and speaker group (based on spoken language or speakers' L1 background), on word duration and vowel distance to the centroid are summarised in Table 4.9.

Table 4.9: Results of statistical analyses concerning the effects of repeated mention (repetition), speaker group based on spoken language or speakers' L1 background, finality and articulation rate.

variable	L1 data subset	English data subset
Whole word duration	Repetition Finality Log(articulation rate) Finality:log(articulation rate)	Repetition Finality Log(articulation rate)
Distance to the centroid	Repetition Log(articulation rate)	Repetition L1

These analyses were carried out using mixed-effects models, and showed that in both subsets, repeated mention significantly improved the model fit for both observed variables. This confirms that the effect of this factor is not merely an artefact of the effect of a word's position as final or non-final in a prosodic unit. In addition to that, a word's position as final or non-final in a prosodic unit proved to have an influence on word durations in both subsets. Words occurring in final position were found to be noticeably longer than words in non-final position, which is in accordance with much previous research (e.g. Oller, 1973; Wightman et al., 1992). Articulation rate was also

Chapter 4

found to significantly contribute to the model fit for word duration in both subsets, and for spectral contrast in the *L1 data* subset. The correlation analyses of this variable with residuals of the mixed model containing the relevant random effects, however, showed only weak ties. Unfortunately, we were not able to come up with a better-suited method to test these relationships in our data. An interesting finding was the significant interaction of the effect of articulation rate and a word's finality status on the word duration in the *L1 data* subset, showing a fair correlation of articulation rate with word duration in non-final items, while duration of words in final positions in a prosodic unit showed no correlation with the local articulation rate.

5 Summary and conclusions

The aim of this thesis was to shed light on the relatively little explored area of phonetic reduction in non-native production. Chapter 1 introduced the topic of the thesis and presented an overview of the relevant literature, focussing in more detail on the area of speaking style research (Section 1.2.1.2), studies investigating repeated mentions of words within a discourse (Section 1.2.1.3) and relevant issues in non-native production research (Section 1.2.2). As presented in Chapter 2, speech material of different types was recorded for the purposes of the present investigation, using both native and non-native speakers with different L1 backgrounds. The investigation itself comprises two studies addressing the research topic from different angles. While Chapter 3 focussed on the realisations of English function words in read and spontaneous speech (produced by native and non-native speakers), Chapter 4 investigated the phonetic reduction of repeated mentions of content words in spontaneous dialogues in Czech and Norwegian spoken by native speakers, as well as the reduction of English content words produced by both native and non-native speakers. Each of the chapters introduced the methods used in the study, presented the results of statistical analyses, and lastly, discussed the results in a comparison with relevant previous research (Sections 3.4 and 4.4, respectively). This final chapter summarises the outcomes of the work and the main findings (Section 5.1). In Section 5.2, general conclusions are presented and related to the hypotheses of the thesis (see Section 1.4.2). The last section (Section 5.3) mentions the identified methodological issues and provides suggestions for future research.

5.1 Summary

5.1.1 Speech material collection

The aim of this thesis was to investigate the patterns of phonetic reduction in spontaneous non-native production. For this purpose, we collected different types of speech material from three groups of speakers with different L1 backgrounds: Czech, English and Norwegian. We recruited 10 speakers per speaker group. In addition to

Chapter 5

recording the Czech and the Norwegian speakers' productions in non-native English, we also obtained recordings in the speakers' L1. Although no proficiency testing was carried out prior to the recording, the speakers (for the L2 recordings) were selected carefully to make sure their proficiency is sufficient for the required task. Subsequently, non-native speaker proficiency was evaluated on randomly selected portions of the spontaneous speech material in L2 using several fluency-related measures. This comparison showed that while there was a noticeable variation in the fluency between the speakers in each group, there were no large systematic differences between the two groups based on the speakers' L1 backgrounds.

The core of the material consists of spontaneous dialogues elicited from pairs of speakers using a conversational task ("picture replication task", for details see Section 2.3.1). One of the motives for using a conversational task was the need to obtain spontaneous speech in the speakers' non-native language. The selected task proved to be suitable for this purpose, as it amused the speakers and stimulated lively conversations. We may assume that concentrating on the task objective distracted the speakers from focussing excessively on their speech production (which was considered particularly important in the non-native production situation). The resulting dialogues lasted for the most part between 30 and 40 minutes, or longer. The total duration of the spontaneous material amounted to over 16 hours.

The other part of the material, used in one of the studies, was read speech in English. This material consisted of BBC news transcripts read by the same speakers that produced the spontaneous dialogues (i.e. native English, Czech and Norwegian speakers).

5.1.2 Realisations of English function words in read and spontaneous speech

This part of the investigation explored the realisations of three English function words: *in*, *of* and *to*, in read and spontaneous speech produced by native and non-native speakers. The study described the temporal organisation as well as spectral properties of

the observed words. In addition, an attempt was made to determine the effects of some aspects of surrounding segmental context (e.g. segment type, phonetic voicing) and articulation rate on the function words under investigation.

The results showed that many of the observed variables describing the realisations of function words were significantly influenced by the speakers' L1 background. The patterns of L1 effects, however, differed between the three function words. While for the preposition *in*, none of the observed variables was found to vary depending on the speakers' L1, in the words *of* and *to*, speakers' L1 background was shown to affect several measures relating to temporal organisation as well as spectral properties. In particular, normalised duration and vowel proportion measure (i.e. relative duration of vowel within the word duration) in both of these function words differed between the groups of speakers with different L1. While in the word *to*, the native speaker group produced significantly shorter durations and lower vowel proportions than both non-native groups, the patterns found in the word *of* were less straightforward. Normalised durations were longer in the Czech speakers' productions than in productions of both the native English and Norwegian speakers, while the vowel proportions were higher in the Norwegian speakers' items, compared to the other two speaker groups. In sum, longer word durations in (some) non-native speakers' productions and non-natives' longer vowel durations (cf. more detailed discussion of the results for the word *of* in Section 3.4.2) seem to indicate the non-natives' insufficient degree of reduction of these two function words in comparison to native productions. Further, a higher proportion of phonetic voicing in the fricative in the preposition *of* was found in Norwegian items, compared to the other two speaker groups. This finding may result from the differences between the phonological systems of the native languages of the three speaker groups, particularly the phonetic properties associated with phonological voicing contrast and phonological processes related to voicing (see Section 3.4.2 for details). Also the plosive release proportion differed depending on the speakers' native language, with native English speakers producing a relatively long release compared to both non-native groups. This may be associated with tendencies to plosive affrication described to occur in casual speech in some dialects of English. Vowel formants were also found to differ

Chapter 5

depending on speaker group. The items produced by native speakers showed more fronted vowel quality in the word *of*, and more open vowel quality in the word *to*. Both of these findings can be seen as tendencies to produce more centralised vowel qualities by the natives as compared to both non-native groups. Lastly, friction intensity in the fricative (for the word *of*) and in the plosive release (for the word *to*) was found to differ between the speaker groups. In the word *of*, native speakers showed higher friction intensity than both non-native groups, and the Norwegians had lower values than the Czech speakers. In the plosive release in the word *to*, friction intensity was also lower in items spoken by the Norwegian speakers. Apart from the apparent relation of friction intensity to voicing proportion in the fricative (for the word *of*), these findings seem to relate to qualities of corresponding sounds in the L1 phonological systems of the three speaker groups.

We also revealed the significant overall effects of speaking style on some of the observed measures, across the speaker groups (native and non-native). Speaking style particularly affected the preposition *of*. The items produced in spontaneous speech had shorter normalised word durations, higher vowel proportions and higher voicing proportions within the fricative. As for the spectral measures, friction intensity in the fricative, a measure strongly (negatively) correlated with voicing proportion, was found to be lower in spontaneous speech. All of these findings may be related to reduction processes causing overall shortening, a tendency to shorten or omit the fricative in some contexts, and fricative lenition. In the word *to*, a higher proportion of plosive release, corresponding to a relatively shorter closure phase, was found in spontaneous items. Moreover, cases where the plosive was realised without a complete closure occurred more frequently in spontaneous speech. Both of these findings indicate a tendency to consonant lenition resulting from a decreased effort during plosive closure articulation in spontaneous speech. Further, speaking style was found to affect the front-back dimension of vowel articulation in the word *to* (as inspected using raw F2 values in Bark, separately for males and females), resulting in more fronted vowel qualities in read speech as compared to those in spontaneous items. This tendency was, however, only significant in the male speaker group. Lastly, we found lower values of friction

intensity in the plosive release of the word *to* in read as compared to spontaneous items. Both of the latter findings seem to be caused by other aspects of speaking style than the varying degree of production effort and precision, and remain to be further investigated. However, this lies beyond the scope of the present research. In addition to the main effects of speaking style, speaking style effect on all of the observed temporal measures, as well as on the friction intensity in the word *of*, was found to interact with the speakers' L1 background. A possible explanation for all of these interactions is the occurrence of unusually long fricatives in the Czech speakers' production in read speech (cf. more detailed discussion of the results for the word *of* in Section 3.4.2).

In the analyses of the effects of certain aspects of context on the realisations of function words, we revealed largely consistent tendencies across the different speaker groups. In the words *of* and *to*, the following segment type (consonant vs. vowel) was found to have an effect on the normalised durations, consistent with previous research, and the type of segment immediately neighbouring with the vowel in the observed word influenced the measure of vowel openness. We also observed a greater degree of nasalisation of the vowel in the word *in* following a (phonologically) nasal sound, and a clear effect of the presence of phonetic voicing in the following segment on voicing proportion and friction intensity in the fricative of the word *of*. This effect was consistent across the speaker groups. We may assume that the observed effects are universal (physiological), provided there is no interference with relevant aspects of the speakers' L1 phonological system.

5.1.3 Reduction of repeated mentions of content words

This part of the investigation inspected the reduction of repeated occurrences of content words produced by the same speaker within a dialogue. The first mention of a word by a speaker was compared with two later mentions of the same word by the same speaker, using productions in Czech, Norwegian, and English produced by native and non-native speakers. The study focussed on durational, rhythmical and spectral aspects of reduction in the observed words. The effects of factors related to prosodic structure, discourse status of the observed words and articulation rate were also taken into consideration.

Chapter 5

The data were analysed in two subsets, separately addressing the material produced in different languages spoken by their native speakers, and English material spoken by native as well as non-native speakers.

The results showed in the first place a consistent significant effect of repeated mention on all durational measures as well as on the measure of spectral contrast (i.e. stressed vowels' distance to the centroid) in all the observed groups of items (produced in different languages spoken by the native speakers, as well as in English spoken by non-native speakers). Repeated mentions of words were generally shorter and contained more centralised vowels in their stressed syllables as compared to the word's first mention. In our additional analyses of the factors related to prosodic structure, we also confirmed that the effect of repeated mention is not an artefact of unevenly distributed word accentuation or position within a prosodic unit. In addition, some of the investigated measures (e.g. syllable duration, spectral contrast) were significantly affected by the language spoken. This finding was, however, not considered relevant to the objectives of the investigation. Even though the detected differences may have been caused by different properties of the respective languages, it was possible that they were just due to the composition of samples for each of the speaker groups. It must be noted that the samples of items were of limited size, and thus relatively susceptible to such bias. Lastly, the rhythm measure, unstressed-to-stressed syllable duration ratio (in polysyllabic words), was generally found to remain constant across the repeated mentions, indicating that any durational differences between first and repeated mentions of words are distributed proportionally across the stressed and unstressed syllables. However, an exception from this tendency was revealed for the items in English produced by the Czech speakers, which showed much greater unstressed-to-stressed syllable duration ratios in their first mentions of words.

Regarding the effects of the observed control factors, we first investigated the effect of the factor relating to the discourse status of the word, dialogue firstness, on three of the measures. While the overall effect of repeated mention on reduction remained significant even after the inclusion of this factor in the analyses, we revealed an

interaction of the factors repeated mention and firstness to affect the observed durational measures (word duration and mean syllable duration) in both data subsets, as well as the spectral contrast measure in the English subset of data. The interaction indicates that durational and spectral reduction occurs particularly in the repeated mentions of triplets, where the first mention by the given speaker was actually the word's very first mention within the dialogue. The non-first triplets, where the word had been previously produced by the other speaker, on the other hand, do not show a significant tendency to reduction. An exception from this pattern was, however, revealed in English spoken by the Czech speakers. It turns out that the Czech speakers' English productions show durational reduction even in items where the first triplet member was not the word's first occurrence in the dialogue. This may indicate the Czech speakers' limited ability to adjust the durational properties in their English productions in relation to the word's overall discourse status, resulting in a more mechanical durational reduction of words once they are produced repeatedly by the same speaker.

The analyses inspecting the effect of the word's position as final or non-final in a prosodic unit confirmed a general tendency to longer durations of words in final positions. No such effect of word position was found on the spectral contrast in the stressed vowel. Moreover, an interaction of finality and articulation rate effects on word duration was found in the material spoken in different languages by their native speakers. A fair correlation of the articulation rate with word duration was found in non-final items, while the duration of words in final position in a prosodic unit showed no significant correlation with the local articulation rate.

5.2 Conclusions

The studies presented in Chapters 3 and 4 have brought diverse findings about the patterns of phonetic reduction in non-native speech. The results were summarised and discussed in detail at the end of each chapter (see Sections 3.4 and 4.4), while the previous section (Section 5.1) provided an overview of the most important findings. The following paragraphs will relate the findings to the hypotheses of this thesis (see Section

1.4.2) and present the general conclusions regarding the tendencies to phonetic reduction in non-native speech as well as some deviating patterns that seem to relate to the non-native speakers' L1, as identified in the present investigation.

5.2.1 Reduction tendencies in non-native production

Firstly, the results show that certain factors known to influence the degree of phonetic reduction in several languages produced by native speakers have a comparable influence on non-native speakers. In particular, a consistent effect of *repeated mention* on durational as well as spectral reduction of the content words in English productions of both non-native speaker groups included in the investigation was revealed by the results of the study described in Chapter 4. We may assume that the universal effect of the word's status as "new" or "given" in the discourse is due to its close association with some fundamental principles of speech communication. Sharing relevant information in an efficient way may be assumed to be a general goal of speech communication regardless of the spoken language (and speakers' L1). Choosing an appropriate phonetic form from a continuum between very clear (hyperarticulated) and extremely reduced (hypoarticulated) forms may have originated as a subconscious mechanism for reducing articulatory effort and thus increasing the efficiency of the communication process. It should be noted that the material used for the present study was apparently well suited for this kind of research. In contrast to that, monologue speech or even read speech may be expected to be less affected by this communication-based factor. In sum, the results of the study, demonstrating the general effect of repeated mention for native as well non-native production, bring evidence supporting the hypothesis H0: the general tendency to phonetic reduction in non-native productions is similar to that of native speakers.

A less clear pattern of results was revealed in the analyses of the effects of *speaking style* on the realisations of English function words by native and non-native speakers, presented in Chapter 3. While a number of observed phonetic measures were found to differ in the two investigated speaking styles, only some of them may be reliably related to the varying degree of production effort and precision. For example, shorter

normalised word durations, a higher proportion of vowel and higher proportion of voicing within the fricative in the preposition *of*, or a relatively shorter duration of plosive closure in the preposition *to* in spontaneous as compared to read speech may be seen as lenitions resulting from a decrease of articulatory effort in spontaneous speech. On the other hand, we found other differences between the speaking styles that are probably related to other aspects of speaking style specification. After all, according to Eskénazy (1993), intended clarity of speech production is only one out of three dimensions that define a given speaking style. We may speculate that some of the characteristics we observed in read speech result to some degree from an unconscious imitation of a “model behaviour” for the given speaking style, namely the distinctive style of professional BBC newsreaders. Another factor that may complicate the formulation of clear conclusions is the previously observed existence of speaker-specific strategies to achieve a given speaking style (Holm, 2001: 50-51).

In addition to exploring the effects of the two above-mentioned factors on phonetic reduction in non-native speech production, our research identified a possible cause of divergence of non-native productions from native-like reduction patterns. As the results of Chapter 3 show, significant differences both with regard to temporal organisation and to spectral properties (particularly vowel formants) were found between native and non-native productions of the *weak form words of* and *to*. In contrast to that, no significant differences between native and non-native realisations were found in the third investigated function word *in*, which does not belong to weak form words. The inadequate non-natives’ realisations of the two weak form words, as opposed to the native-like pattern in the word *in*, may then be explained by the non-native speakers’ insufficient awareness of the weak forms. Native-like use of weak forms is then apparently related to patterns of language use that are language specific, and that non-native speakers lacking extensive experience in an L2 environment have a small chance to master. Apart from the theoretical impact of this finding, it may be useful for L2 teaching. For example, specific training aimed at the use of weak forms of function words may be expected to improve the native-likeness of L2 production at any stage of the L2 learning process. This point also presents concrete evidence of an insufficient

Chapter 5

degree of reduction in non-native productions in certain conditions, thus supporting the hypothesis HA1 of this thesis. We must admit, however, that the present investigation did not address the question of the effect of the amount of the speakers' L2 experience on the degree of reduction in their L2 production. The only relevant findings were the significant correlations of some of the measures of speech fluency (in speakers' L2) with rough measures relating to the speakers' L2 experience (for details see Section 2.2.2). For a thorough study of the relationship between L2 experience and degree of reduction in L2 speech production, a larger number of speakers with different amounts of L2 experience would be needed.

Overall, our investigation supports both of the opposing hypotheses H0 and HA1. On the one hand, the degree of phonetic reduction seems to vary in response to certain factors related to predictability, in a similar manner for native and non-native speakers. This general, language-independent tendency is consistent with the predictions of the H&H theory (Lindblom, 1990). On the other hand, non-native speakers in this study did not reduce weak form words sufficiently. This suggests that certain patterns of phonetic reduction are language-specific. These seemingly contradictory results are in agreement with findings of related research on clear speech strategies. For instance, late Finnish-English bilinguals in the study by Granlund et al. (2012) were found to produce global clear speech modifications in a native-like manner, but they applied the same strategies to enhance some of the segmental contrasts both in their L1 and L2, without regard to language-specific cue-weighting.

5.2.2 Deviating patterns

An interesting pattern of results that seems to relate to non-native speakers' L1 background was revealed in the analysis of rhythmical aspects of reduction of repeated mentions of polysyllabic content words. While the overall values of the rhythm measure expressing the ratio of unstressed to stressed syllable durations could not be used to directly compare speaker groups with different L1 backgrounds (due to the fact that the studied samples were composed of different lexical items), a significant interaction of repeated mention and speakers' L1 background was revealed. This interaction was due

to Czech speakers' unusually long unstressed syllables and resulting higher unstressed-to-stressed syllable duration ratios occurring in the first mentions of words. In contrast to that, the ratio of unstressed and stressed syllables did not change much between the first and the two repeated mentions in the productions of natives, as well as Norwegian speakers of English. We may speculate whether this peculiar rhythmical form of hyperarticulated polysyllabic words is a result of Czech speakers' particularly hesitant production when mentioning a given word for the first time. However, considering the substantial differences in phonological properties related to rhythm type between Czech on one side, and English and Norwegian on the other (for more details about rhythm type and related phonological properties of the three languages see Section 1.3.1), it seems plausible that the potential for such deviation from the expected pattern of the target language (English) is due to the properties of the speakers' L1 (Czech).

In addition, in the study of the realisations of English function words by native and non-native speakers, we identified several details where properties of the speakers' L1 could explain the differences between the non-native realisations and the native production pattern. Here we can mention the Norwegian speakers' higher fricative voicing proportion in the preposition *of*, as compared to the other two speaker groups. This pattern of results may be explained by the phonetic properties associated with phonological voicing contrast and phonological processes related to voicing in the native languages of the three speaker groups (see Section 1.3.3). Further, the occurrence of more fronted vowels (i.e. higher F2) in the function word *to* produced by Norwegian males may be explained by Norwegian speakers' tendency to substitute English sounds /ʊ/ and /u/ with Norwegian vowel /ʉ/ (e.g. Davidsen-Nielsen 1977: 92-93, 98). The Czech speakers' lower degree of nasalisation in the segment following the preposition *in* may also be put in relation with certain regularities of the speakers' L1 (Czech), namely the tendency to indicate word boundaries using different means. For example, while within words, assimilation of place of articulation as well as some consonant elisions occur regularly and are in some cases even obligatory according to Czech orthoepic rules, the occurrence of such phenomena across a word boundary is considered typical for substandard production (Palková, 1997: 327-328, 332-336).

Similarly, the glottalisation of word-initial vowels is very frequent in Czech, being described as obligatory in some contexts in the Czech pronunciation norm (Palková, 1997: 325-327). The absence of glottalisation is then considered a sign of careless speech. Furthermore, it was shown that the glottalisation of word-initial vowels occurred frequently in the non-native English productions of Czech speakers (Bissiri and Volín, 2010). Although the particular phenomena mentioned in this paragraph may appear less generalisable since they only refer to phenomena relating to the realisations of the particular function words, they still seem to illustrate the assumption that the characteristics of a speakers' L1 may relate to deviations in their productions from the native-like patterns. In sum, the phenomena mentioned in this section may be considered examples corroborating the hypothesis HA2 of this thesis.

5.3 Methodological issues and suggestions

Apart from the main results summarised in the previous sections, this thesis has gathered a number of relevant practical observations that refer to the material and methods of the investigation and illustrate the challenges faced in spontaneous speech research. Some of these observations may be useful in designing future experiments using similar types of material.

In Chapter 2, a detailed description of the materials used throughout this thesis is provided. There is no doubt that obtaining speech material from non-native speakers is a challenging task, even more so when spontaneous speech is required. We observed that in spite of a relatively careful selection of speakers, a large variability in speaker fluency (possibly partly resulting from the amount of L2 experience) has to be expected. Speaker proficiency tests administered prior to the recording may provide more fine-grained information, and thus facilitate the selection of suitable speakers. However, proficiency testing in itself is an intricate task, and the results typically refer to the speakers' grammatical or lexical knowledge. Therefore, for the purposes of the present research, proficiency testing was not considered to be appropriate. Similarly, a subsequent selection of the speakers for an investigation from a larger set of recordings

may be a good solution to reduce the undesired variation in speaker proficiency level. Such a thorough approach is usually only feasible in large research projects.

Moreover, as many of the analyses showed, the amount of material in the presented studies was in some cases not sufficient to carry out detailed analyses including various context-related factors. The multitude of co-occurring factors on all levels of description and the practical impossibility of controlling for all of them are an inherent drawback of spontaneous speech research, contrasting with the relative simplicity of well-designed experiments using laboratory materials. A solution suitable for some types of research questions would be the use of large corpora of automatically annotated material. However, to guarantee a high degree of accuracy in annotations, or to obtain more detailed descriptions, the research still needs to rely on skilled human annotators.

An issue encountered during the analyses of vowel formants in Chapter 3 resulted in a practical suggestion with regard to methods for formant analysis. To describe vowel quality in the observed items, we intended to use the Bark distance normalisation of formant values, which according to previous research reduces the gender-related variation in the formant values. In the course of the analyses, however, it turned out that in our data especially the F1 - F0 measure still retains a significant between-gender variation. This urged us to perform additional analyses of raw formant values in Bark (separately for males and females). Splitting the analyses into subsets by gender then led to the discovery of partly diverging tendencies, reducing the generality of the findings. In general, this issue suggests that the degree to which Bark distance normalisation reduces anatomical variation may vary with the types of material used. Therefore, attention should be paid to checking normalisation characteristics on specific material.

Our treatment of the segmental context effects and the influence of some aspects of prosodic structure has had clear limitations. For instance, various other aspects of segmental context could have been classified, such as place and manner of articulation of neighbouring sounds. However, the amount of studied items per speaker would have

to be increased accordingly to avoid strongly unbalanced designs. Such material would then enable a more precise evaluation of the coarticulatory tendencies in the observed items. Similarly, a more fine-grained description of various aspects of prosodic structure could help reveal interesting tendencies in the use of prosody by native and non-native speakers and their interactions with other factors. We have to bear in mind, however, that prosodic annotation of spontaneous speech material, and particularly in material spoken by non-native speakers, is an extremely demanding task, and would require very experienced annotators.

5.3.1 Future research directions

In today's globalised world it is beyond doubt that research of various aspects of non-native speech production will remain a relevant and attractive area. Spontaneous speech, as the single most common and natural speaking style, should stay in the centre of attention of such research to enrich the current knowledge of non-native production regularities, which is largely based on research using laboratory speech material.

In general, larger studies are needed, both in terms of the number of involved non-native speakers, and in terms of overall size of the material. Only in sufficiently large corpora can we assess the effects of factors relating to speaker proficiency and experience reliably. It may also be possible to investigate concrete phenomena in the speech production of subjects with corresponding proficiency, comparing the effect of the speakers' L1 background. Collecting speech produced using different elicitation tasks or under different conditions (e.g. time pressure or high cognitive load vs. relaxed chat) may also offer intriguing possibilities to compare the changes of different aspects of speaker proficiency (e.g. segmental accuracy, prosody, fluency measures, etc.) in different situations.

Similarly, substantially larger amounts of material per speaker (in a given speech condition) would be needed to untangle the effects of various aspects of context, prosodic structure and other factors. As mentioned previously, detailed analyses of the effects of place and manner of articulation of neighbouring segments may bring

valuable information about the coarticulation tendencies in non-native productions and the possible effects of the speakers' L1 background.

This research, being an investigation of phonetic reduction in non-native production, did not aim to address any issues relating to the perception of the studied phenomena in non-native speech. We are convinced, however, that an investigation of the perceptual relevance of varying degree of reduction in non-native speech may bring very interesting findings. For example, exploring the effect of varying degree of reduction in non-native speech on the degree of perceived foreign accent or on speech intelligibility would be a valuable contribution to the existing research in these areas. In particular, the findings from such research may be relevant for identifying concrete issues in L2 production that deserve increased attention (in L2 instruction). We hope that the results of future research will not only further advance the understanding of the regularities of L2 speech acquisition, but also contribute to the area of L2 instruction.

6 References

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Adank, P., Smits, R. and van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116 (5), pp. 3099-3107.
- Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34 (4), pp. 351-366.
- Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47 (1), pp. 31-56.
- Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J. and Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, 39 (1), pp. 1-17.
- Baker, R. E. and Bradlow, A. R. (2007). Second mention reduction in Indian English and Korean (A). *Journal of the Acoustical Society of America*, 122 (5), p. 2993.
- Baker, R. E. and Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, 52 (4), pp. 391-413.
- Baker, W. and Trofimovich, P. (2005). Interaction of native- and second-language vowel systems(s) in early and late bilinguals. *Language and Speech*, 48 (1), pp. 1-27.
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G. and Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42, pp. 1-22.
- Barry, M. C. (1992). Palatalisation, assimilation and gestural weakening in connected speech. *Speech Communication*, 11, pp. 393-400.
- Barry, W. J. (1995). Phonetics and phonology of speaking styles. In: *Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm*, vol. 2, pp. 4-10.
- Barry, W. and Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, 31 (1), pp. 51-66.
- Beckman, M. and Ayers Elam, G. (1997) *Guidelines for ToBI Labeling, version 3.0*. Unpublished manuscript, Ohio State University.
- Bell, A., Gregory, M. L., Brenier, J. M., Jurafsky, D., Ikeno, A. and Girand, C. (2002). Which predictability measures affect content word durations? In: *ISCA Tutorial and*

Research Workshop (ITRW) on Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C. and Gildea, D. (1999). Forms of English function words--effects of disfluencies, turn position, age and sex, and predictability. In: *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1-7 August 1999*.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M. and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, 113 (2), pp. 1001-1024.

Benus, S., Gravano, A. and Hirschberg, J. (2007). The prosody of backchannels in American English. In: *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 6-10 August 2007*, pp. 1065-1068.

Bissiri, M. P. and Volín, J. (2010). Prosodic structure as a predictor of glottal stops before word-initial vowels in Czech English. In: *Proceedings of the 20th Czech-German Workshop – Speech Processing, Prague, September 2010*, pp. 23-28.

Boersma, P. and Weenink, D. (2009). *Praat: Doing phonetics by computer (version 5.1.09)*. [Software]. Retrieved July 7, 2009, from <http://www.praat.org/>.

Bohn, O.-S. and Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 14, pp. 131-158.

Bolotova, O. (2003). On some acoustic features of spontaneous speech and reading in Russian (quantitative and qualitative comparison methods). In: *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 3-9 August 2003*, vol 1, pp. 913-916.

Bond, Z. S. and Fokes, J. (1985). Non-native patterns of English syllable timing. *Journal of Phonetics*, 13, pp. 407-420.

Bondarko, L. V., Volskaya, N. B., Tananaiko, S. O. and Vasilieva, L. A. (2003). Phonetic properties of Russian spontaneous speech. In: *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 3-9 August 2003*, vol 3, pp. 2973-2976.

Bradlow, A. R., Ackerman, L., Burchfield, L. A., Hesterberg, L., Luque, J. and Mok, K. (2011). Language- and talker-dependent variation in global features of native and non-native speech. In: *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, 17-21 August 2011*, pp. 356-359.

Broch, O. (1935). Rhythm in the spoken Norwegian language. *Transactions of the Philological Society*, 34 (1), pp. 80-112.

Browman, C. P. and Goldstein, L. (1992). "Targetless schwa": an articulatory analysis. In: Docherty, G. J. and Ladd, D. R. (eds.) *Papers in laboratory phonology II: Gesture, segment, prosody*, pp. 26-56.

- Butschkow, R. (2002). *Da stimmt doch was nicht!* Frankfurt am Main: Baumhaus Verlag GmbH.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, pp. 39-54.
- Campbell, N. (2004). Speech & Expression; the value of a longitudinal corpus. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC), Lisbon, Portugal*, pp. 183-186.
- Cornejo, R. J., Weinstein, A. C. and Najar, C. (1983). *Eliciting Spontaneous Speech in Bilingual Students: Methods and Techniques*. [Online] ERIC - Education Resources Information Center. Available at: <http://www.eric.ed.gov/PDFS/ED230351.pdf> [Accessed 28 June 2010].
- Cucchiari, C., Strik, H. and Boves, L. (2002). Quantitative assessment of second language learners' fluency: comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, 111 (6), pp. 2862-2873.
- Cutler, A., Garcia Lecumberri, M. L. and Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: effects of local context. *Journal of the Acoustical Society of America*, 124 (2), pp. 1264-1268.
- Dankovičová, J. and Dellwo, W. (2007). Czech speech rhythm and the rhythm class hypothesis. In: *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 6-10 August 2007*, pp. 1241-1244.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, pp. 51-62.
- Daividsen-Nielsen, N. (1977). *English Phonetics*. Translated from Danish by Barbara Bird and Per Moen. Oslo: Gyldendal Norsk Forlag.
- de Silva, V., Iivonen, A., Bondarko, L. V. and Pols, L. C. W. (2003). Common and language dependent phonetic differences between read and spontaneous speech in Russian, Finnish and Dutch. In: *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 3-9 August 2003*, vol 3, pp. 2977-2980.
- Deterding, D. (1997). The formants of monophthong vowels in standard southern British English pronunciation. *Journal of the International Phonetic Association*, 27, pp. 47-55.
- Eskénazi, M. (1993). Trends in speaking styles research. In: *Third European Conference on Speech Communication and Technology, Berlin, Germany, 22-25 September 1993*, pp. 501-509.
- Face, T. L. (2003). Intonation in Spanish declaratives: differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics*, 2, pp. 115-131.
- Fintoft, K. (1961). The duration of some Norwegian speech sounds. *Phonetica*, 7, pp. 19-39.

- Fisher, C. and Tokura, H. (1995). The given-new contract in speech to infants. *Journal of Memory and Language*, 34, pp. 287-310.
- Flege, J. E. (1995). Second-language speech learning: theory, findings and problems. In: Strange, W. (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium: York Press, pp. 233-277.
- Flege, J. E. and Bohn, O.-S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition*, 11 (1), pp. 35-62.
- Flege, J. E., Bohn, O.-S. and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, pp. 437-470.
- Flege, J. E. and MacKay, I. R. A. (2010). "Age" effect on second language acquisition. In: *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010, Poznań, Poland, 1-3 May 2010*, pp. 113-118.
- Flege, J. E., Munro, M. J. and MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97 (5), pp. 3125-3134.
- Florentine, M., Buus, S., Scharf, B. and Canevet, G. (1984). Speech reception thresholds in noise for native and non-native listeners. *Journal of the Acoustical Society of America*, 75 (Suppl. 1), p. S84.
- Fosler-Lussier, E. and Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication*, 29, pp. 137-158.
- Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, 31 (4), pp. 307-319.
- Fowler, C. A. and Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26, pp. 489-504.
- Fowler, C. A., Levy, E. T. and Brown, J. M. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language*, 37, pp. 24-40.
- Fox Tree, J. E. and Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62, pp. 151-167.
- Furui, S., Nakamura, M., Ichiba, T. and Iwano, K. (2005). Analysis and recognition of spontaneous speech using *Corpus of Spontaneous Japanese*. *Speech Communication*, 47, pp. 208-219.
- Garcia Lecumberri, M. L. and Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America*, 119 (4), pp. 2445-2454.

- Gick, B., Wilson, I., Koch, K. and Cook, C. (2004). Language-specific articulatory settings: evidence from inter-utterance rest position. *Phonetica*, 61, pp. 22-233.
- Gimson, A. C. (1980). *An Introduction to the Pronunciation of English*. 3rd ed. London: Edward Arnold.
- Godfrey, J. J., Holliman, E. C. and McDaniel, J. (1992). SWITCHBOARD: telephone speech corpus for research and development. In: *Proceedings of International Conference on Acoustics Speech and Signal Processing ICASSP-92 IEEE*, pp. 517-520.
- Grabe, E. and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In: Warner, N. and Gussenhoven, C. (eds.) *Laboratory phonology 7*. Berlin: Mouton de Gruyter, pp. 515-546.
- Granlund, S., Hazan, V. and Baker, R. (2012). An acoustic-phonetic comparison of the clear speaking styles of Finnish-English late bilinguals. *Journal of Phonetics*, 40, pp. 509-520.
- Green, G. (2004). "Listening makes my brain hurt": issues of listening for students who are speakers of languages other than English. *Directions: Journal of Educational Studies*, 26 (2), pp. 57-71.
- Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E. and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In: *CLS 35, 22-24 April 1999*, pp. 151-166.
- Grenon, I., White, L. (2008). Acquiring rhythm: a comparison of L1 and L2 speakers of Canadian English and Japanese. In: *Proceedings of the 32nd Boston University Conference on Language Development (BUCLD)*, pp. 155-166.
- Guion, S. G. (1995). Word frequency effects among homonyms. *Texas Linguistic Forum*, 35, pp. 103-116.
- Guion, S. G., Flege, J. E., Liu, S. H. and Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, 21, pp. 205-228.
- Gut, U. (2003). Non-native speech rhythm in German. In: *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 3-9 August 2003*, pp. 2437-2440.
- Gut, U. (2007). Learner corpora in second language prosody research. In: Trouvain, J. and Gut, U. (eds.) *Non-native Prosody: Phonetic Description and Teaching Practice*. Berlin/New York: Mouton de Gruyter, pp. 145-167.
- Halle, M., Hughes, G. W. and Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America*, 29 (1), pp. 107-116.
- Harmegnies, B. and Poch-Olivé, D. (1992). A study of style-induced vowel variability: laboratory versus spontaneous speech in Spanish. *Speech Communication*, 11, pp. 429-437.

- Hawkins, S. and Warren, P. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*, 22, 493-511.
- Hazan, V. and Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America*, 130 (4), pp. 2139-2152.
- Hedbávná, B. (2004). Variability of formant structure of Czech vowels in connected speech. *Acta Universitatis Carolinae – Philologica*, 1, pp. 19-31.
- Hieke, A. E. (1984). Linking as a marker of fluent speech. *Language and Speech*, 27 (4), pp. 343–354.
- Hirschberg, J. (2000). A corpus-based approach to the study of speaking style. In: Horne, M. (ed.) *Prosody: Theory and Experiment. Studies presented to Gösta Bruce*. Dordrecht: Kluwer Academic Publishers, pp. 335-350.
- Holm, S. (2001). *Individuell bruk av akustiske parametre ved veksling mellom lest og spontan tale*. Master's thesis, Trondheim: NTNU.
- Holm, S. (2008). *Intonational and Durational Contributions to the Perception of Foreign-accented Norwegian: An Experimental Phonetic Investigation*. Unpublished PhD dissertation. Trondheim: NTNU.
- House, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, 33, pp. 1174-1178.
- House, A. S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 25 (1), pp. 105-113.
- Howell, P. and Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, 10 (2), pp. 163-169.
- Husby, O. and Kløve M. H. (2001). *Andrespråksfonologi*. Oslo: Gyldendal Norsk Forlag AS.
- Iacobucci, D. (1995). Analysis of variance for unbalanced data. In: Stewart, D. W. and Vilcassim, N. J. (eds.) *AMA Winter Educators' Conference: Marketing Theory and Practice*, pp. 337-343.
- Ito, K. and Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. In: Sudhoff, S. et al. (eds.) *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter, pp. 229-257.
- Jansen, W. (2007). Phonological 'voicing', phonetic voicing, and assimilation in English. *Language Sciences*, 29, pp. 270-293.
- Jensen, J. T. (1993). *English Phonology*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Johnson, K. (2004). Massive reduction in conversational American English. In: *Spontaneous Speech: Data and Analysis. Proceedings of the 1st session of the 10th international symposium*. Tokyo, Japan: The National International Institute for Japanese Language, pp. 29-54.
- Jones, D., Roach, P., Hartman, J. and Setter, J. (2003). *English pronouncing dictionary*. 16th ed. Cambridge: Cambridge University Press.
- Jones, M. J. and Llamas, C. (2003). Fricated pre-aspirated /t/ in Middlesbrough English: an acoustic study. In: *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 3-9 August 2003*, pp. 655-658.
- Jorschick, A. (2009). *Sound to Sense corpora*. [Online] Sound to Sense. Available at: http://sound2sense.eu/images/uploads/S2S_corpora_june10.pdf [Accessed 28 June 2010].
- Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C. and Raymond, W. (1998). Reduction of English function words in Switchboard. In: *Proceedings of ICSLP-98, Sydney*, pp. 3111-3114.
- Kluender, K. R., Diehl, R. L. and Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *Journal of Phonetics*, 16, pp. 153-169.
- Kohler, K. J. (1990). Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In: Hardcastle, W. J. and Marchal, A. (eds.) *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, pp. 69-92.
- Kohler, K. J. (1998). The disappearance of words in connected speech. *ZAS Papers in Linguistics*, 11, pp. 21-33.
- Kondo, Y. (1994). Targetless schwa: is that how we get the impression of stress-timing in English? In: *Proceedings of the Edinburgh Linguistics Department Conference*, pp. 63-76.
- Koopmans-van Beinum, F. J. (1980). *Vowel contrast reduction: an acoustic and perceptual study of Dutch vowels in various speech conditions*. Amsterdam: Academische Pers B.V.
- Koopmans-Van Beinum, F. J. (1992). The role of focus words in natural and in synthetic continuous speech: acoustic aspects. *Speech Communication*, 11 (4-5), pp. 439-452.
- Koopmans-van Beinum, F. J. and van Bergem, D. R. (1989). The role of 'given' and 'new' in the production and perception of vowel contrasts in read text and in spontaneous speech. In: *Proceedings of the First European Conference on Speech Communication and Technology, Eurospeech '89, Paris, France, 27-29 September 1989*, vol. 2, pp. 2113-2116.
- Kormos, J. and Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32 (2), pp. 145-164.

- Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford: Oxford University Press.
- Laan, G. P. M. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, 22, pp. 43-65.
- Lass, R. (1984). *Phonology: an introduction to basic concepts*. Cambridge: Cambridge University Press.
- Lavoie, L. (2002). Some influences on the realization of *for* and *four* in American English. *Journal of the International Phonetic Association*, 32 (2), pp. 175-202.
- Lee, B., Guion, S. G. and Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition*, 28 (3), pp. 487–513.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, pp. 2018-2024.
- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6 (3), pp. 172-187.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, pp. 1773-1781.
- Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W. J. and Marchal, A. (eds.) *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, pp. 403-439.
- Lindblom, B. (1996). Role of articulation in speech perception: clues from production. *Journal of the Acoustical Society of America*, 99 (3), pp. 1683-1692.
- Llisterri, J. (1992). Speaking styles in speech research. In: *ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language, Dublin, Ireland, 15-17 July 1992*.
- Local, J. (2003). Phonetics and talk-in-interaction. In: *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 3-9 August 2003*, vol. 1, pp. 115-118.
- Machač, P. and Skarnitzl, R. (2007). Temporal compensation in Czech? In: *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 6-10 August 2007*, pp 537-540.
- Machač, P. and Skarnitzl, R. (2009a). *Principles of Phonetic Segmentation*. Praha: Epocha Publishing House.
- Machač, P. and Skarnitzl, R. (2009b). Phonetic analysis of parasitic speech sounds. In: *Proceedings of the 19th Czech-German Workshop on Speech Processing, Prague*, pp. 61-68.

- Matoušek, J., Skarnitzl, R., Machač, P. and Trmal, J. (2009). Identification and automatic detection of parasitic speech sounds. In: *Proceedings of Interspeech, Brighton, United Kingdom, 6-10 September 2009*, pp. 876-879.
- MacKay, I. R. A. and Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: the role of suppression. *Applied Psycholinguistics*, 25, pp. 373–396.
- Manuel, S. Y. (1995). Speakers nasalise /ð/ after /n/, but listeners still hear /ð/. *Journal of Phonetics*, 23, pp. 453-476.
- Mayo, L. H., Florentine, M. and Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40, pp. 686-693.
- McAllister, J., Potts, A., Mason, K. and Marchant, G. (1994). Word duration in monologue and dialogue speech. *Language and Speech*, 37 (4), pp. 393-405.
- Mixdorff, H. and Pfitzinger, H. R. (2005). Analysing fundamental frequency contours and local speech rate in Map Task dialogs. *Speech Communication*, 46, pp. 310-325.
- Moen, I. and Kristoffersen, G. (2006). Norwegian Prosody. *Symposium on Nordic and Baltic Prosody, Copenhagen, 16-17 October 2006*.
- Moon, S.-J. and Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, 96 (1), pp. 40-55.
- Mooshammer, C. and Geng, C. (2008). Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association*, 38 (2), pp. 117-136.
- Morland, A. (2010). *Norwegians' production and perception of the voiceless-voiced contrast in English*. Master's thesis. Trondheim: NTNU.
- Nábělek, A. K. and Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *Journal of the Acoustical Society of America*, 75 (2), pp. 632-634.
- Nakajima, S. and Allen, J. F. (1993). A Study on Prosody and Discourse Structure in Cooperative Dialogues. *Phonetica*, 50 (3), pp. 197-210.
- Nakamura, M., Iwano, K. and Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language*, 22, pp. 171-184.
- Nguyen, T. and Ingram, J. (2004). A corpus-based analysis of transfer effects and connected speech processes in Vietnamese English. In: *Proceedings of the 10th Australian International Conference on Speech Science & Technology, Sydney, 8-10 December 2004*, pp. 516-521.

- Nguyen, T. A. T. and Ingram, J. (2005). Vietnamese acquisition of English word stress. *TESOL Quarterly*, 39 (2), pp. 309-319.
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54 (5), pp. 1235-1247.
- Palková, Z. (1997). *Fonetika a fonologie češtiny*. Praha: Karolinum.
- Palková, Z. and Volín, J. (2003). The role of F0 contours in determining foot boundaries in Czech. In: *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 3-9 August 2003*, vol. 1, pp. 1783-1786.
- Piske, T., MacKay, I. R. A. and Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29, pp. 191-215.
- Pluymaekers, M., Ernestus, M. and Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62, pp. 146-159.
- Podlipský, V. J. and Chládková, K. (2007). Vowel duration variation induced by coda consonant voicing and perceptual short/long vowel categorization in Czech. In: *Proceedings of the 17th Czech-German Workshop – Speech Processing, Prague*, pp. 68-74.
- Poesio, M. and Vieira, R. (1998). A corpus-based investigation of definite description use. *Computational Linguistics*, 24 (2), pp. 183-216.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. [Software]. Vienna: R Foundation for Statistical Computing.
- Ramus, F., Nespore, M. and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, pp. 265-292.
- Riggenbach, H. (1991). Toward an understanding of fluency: a microanalysis of nonnative speaker conversations. *Discourse Processes*, 14, pp. 423-441.
- Roach, P. (1982). On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. In: Crystal, D. (ed.) *Linguistic Controversies*. London: Edward Arnold, pp. 73-79.
- Roach, P. (1983). *English Phonetics and Phonology: A practical course*. Cambridge: Cambridge University Press.
- Rouas, J.-L., Beppu, M. and Adda-Decker, M. (2010). Comparison of spectral properties of read, prepared and casual speech in French. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Valletta, Malta*, pp. 606-611.
- Sangster, C. M. (2001). Lenition of alveolar stops in Liverpool English. *Journal of Sociolinguistics*, 5 (3), pp. 401-412.

- Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L. and Dmitrieva, O. (2007). An acoustic study of real and imagined foreigner-directed speech. In: *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 6-10 August 2007*, pp. 2165-2168.
- Schuppler, B., van Dommelen, W. A., Koreman, J. and Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, 40, pp. 595-607.
- Shields, L. W. and Balota, D. A. (1991). Repetition and associative context effects in speech production. *Language and Speech*, 34 (1), pp. 47-55.
- Shockey, L. (2003). *Sound Patterns of Spoken English*. Malden: Blackwell Publishing.
- Sikveland, R. O., Öttl, A., Amdal, I., Ernestus, M., Svendsen, T. and Edlund, J. (2010). Spontal-N: a corpus of interactional spoken Norwegian. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Valletta, Malta*, pp. 2986-2991.
- Skarnitzl, R. (2011). *Znělostní kontrast nejen v češtině*. Praha: EPOCH.
- Skarnitzl, R. and Machač, P. (2012). Míra rušivosti parazitních zvuků v řeči mediálních mluvčích. *Naše řeč*, 1, pp. 3-14.
- Skarnitzl, R. and Volín, J. (2005). Czech voiced labiodental continuant discrimination from basic acoustic data. In: *Proceedings of Interspeech, Lisbon, 4-8 September 2005*, pp. 2921-2924.
- Smith, C. L. (1997). The devoicing of /z/ in American English: effects of locan and prosodic context. *Journal of Phonetics*, 25, pp. 471-500.
- Spilková, H., Brenner, D., Öttl, A., Vondříčka, P., van Dommelen, W. A. and Ernestus, M. (2010). The Kachna L1/L2 Picture Replication Corpus. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Valletta, Malta*, pp. 2432-2436.
- SPSS Inc. (2006). *SPSS for Windows (version 15.0)*. [Software]. Chicago: SPSS Inc.
- Stevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, 3, pp. 32-49.
- Swerts, M. and Collier, R. (1992). On the controlled elicitation of spontaneous speech. *Speech Communication*, 11, pp. 463-468.
- Syrdal, A. K. (1985). Aspects of a model of the auditory representation of American English vowels. *Speech Communication*, 44, pp. 121-135.
- Syrdal, A. K. and Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79 (4), pp. 1086-1100.

Terken, J. and Hirschberg, J. (1994). Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position. *Language and Speech*, 37 (2), pp. 125-145.

Toivola, M., Lennes, M., Korvala, J. and Aho, E. (2010). A longitudinal study of speech rate and pauses in non-native Finnish. In: *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010, Poznań, Poland, 1-3 May 2010*, pp. 499-504.

Torreira, F., Adda-Decker, M. and Ernestus, M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, 52 (3), pp. 201-212.

Torreira, F. and Ernestus, M. (2009). Probabilistic effects on French [t] duration. In: *Proceedings of Interspeech, Brighton, United Kingdom, 6-10 September 2009*, pp. 448-451.

Towell, R., Hawkins, R. and Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17 (1), pp. 84-119.

Trautmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, 88 (1), pp. 97-100.

Trofimovich, P. and Baker, W. (2006). Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28 (1), pp. 1-30.

Trón, V. (2008). On the durational reduction of repeated mentions: recency and speaker effects. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Marrakech, Morocco, 26 May - 1 June 2008*, pp. 2777-2870.

Uther, M., Knoll, M. A. and Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, 49, pp. 2-7.

Van Bael, C., Baayen, H. and Strik, H. (2007). Segment deletion in spontaneous speech: a corpus study using mixed effects models with crossed random effects. In: *Proceedings of Interspeech, Antwerp, Belgium*, pp. 2741-2744.

van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*, 12, pp. 1-23.

van Dommelen, W. A. (2007). Temporal patterns in Norwegian as L2. In: Trouvain, J. and Gut, U. (eds.) *Non-native Prosody: Phonetic Description and Teaching Practice*. Berlin/New York: Mouton de Gruyter, pp. 121-143.

van Dommelen, W. (2011). *Norwegian formant values*. [e-mail] (Personal communication, 19 April 2011).

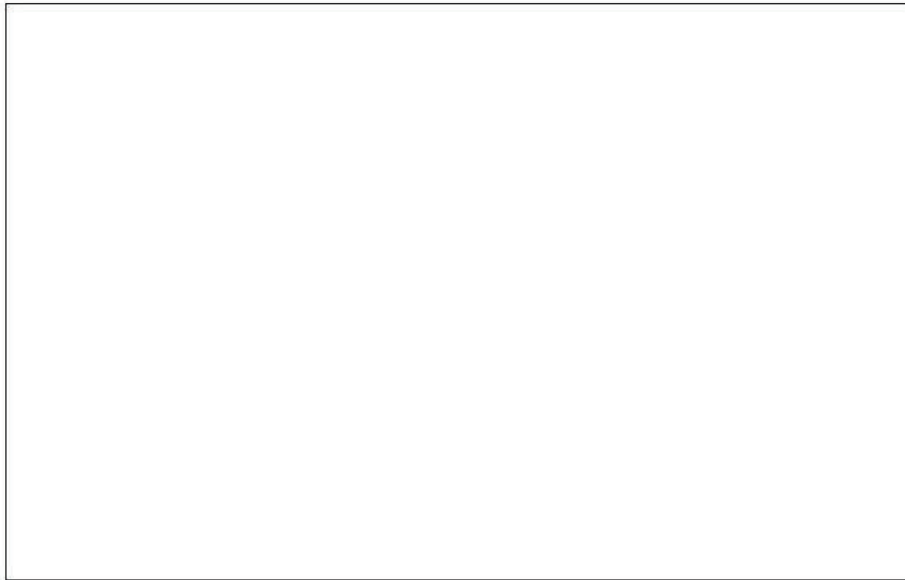
van Dommelen, W. A. and Ringen, C. (2007). Intervocalic fortis and lenis stops in a Norwegian dialect. In: *Proceedings of Fonetik 2007*, pp. 5-8.

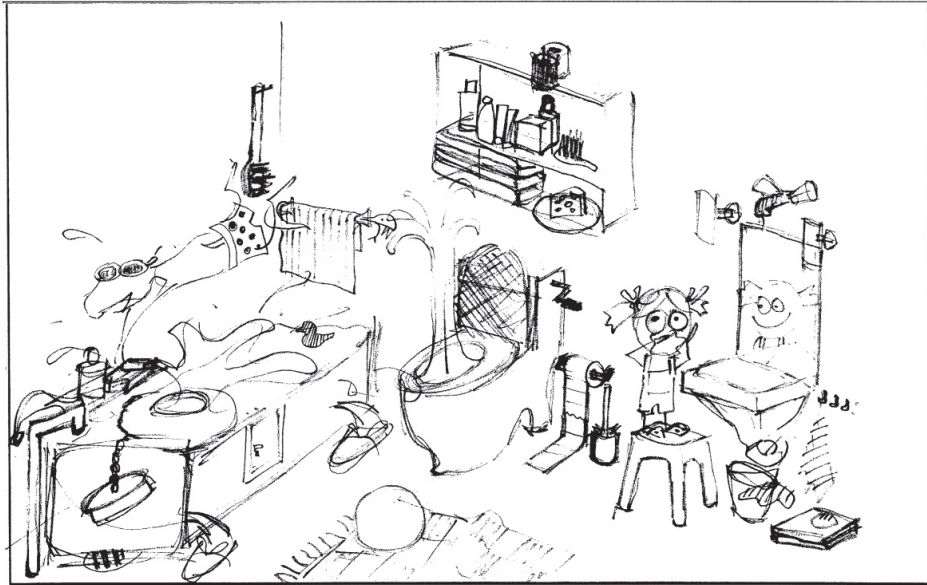
- van Santen, J. P. H. (1992). Contextual effects on vowel duration. *Speech Communication*, 11, pp. 513-546.
- van Son, R. J. J. H. and Pols, L. C. W. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 88 (4), pp. 1683-1693.
- van Son, R. J. J. H. and Pols, L. C. W. (1992). Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 92 (1), pp. 121-127.
- van Son, R. J. J. H. and Pols, L. C. W. (1999). Acoustic description of consonant reduction. *Speech Communication*, 28, pp. 125-140.
- van Wijngaarden, S. J., Steeneken, H. J. M. and Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America*, 111 (4), pp. 1906-1916.
- Volín, J. (2005). Rhythmical properties of polysyllabic words in British and Czech English. In: Čermák, J., Klégr, A., Malá, M. and Šaldová, P. (eds.) *Patterns: a festschrift for Libuše Dušková*. Praha: Kruh moderních filologů, pp. 279-292.
- Weber, A. and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, pp. 1-25.
- Wenk, B. J. (1985). Speech rhythms in second language acquisition. *Language and Speech*, 28 (2), pp. 157-175.
- White, L. and Mattys, S. L. (2007). Calibrating rhythm: first language and second language studies. *Journal of Phonetics*, 35, pp. 501-522.
- White, L., Wiget, L., Rauch, O. and Mattys, S. L. (2010). Segmentation cues in spontaneous and read speech. In: *Proceedings of the Fifth Conference on Speech Prosody, Chicago, 11-14 May 2010*, pp. 1-4.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91 (3), pp. 1707-1717.
- Wilson, I. L. (2006). *Articulatory settings of French and English monolingual and bilingual speakers*. Unpublished PhD dissertation, University of British Columbia.

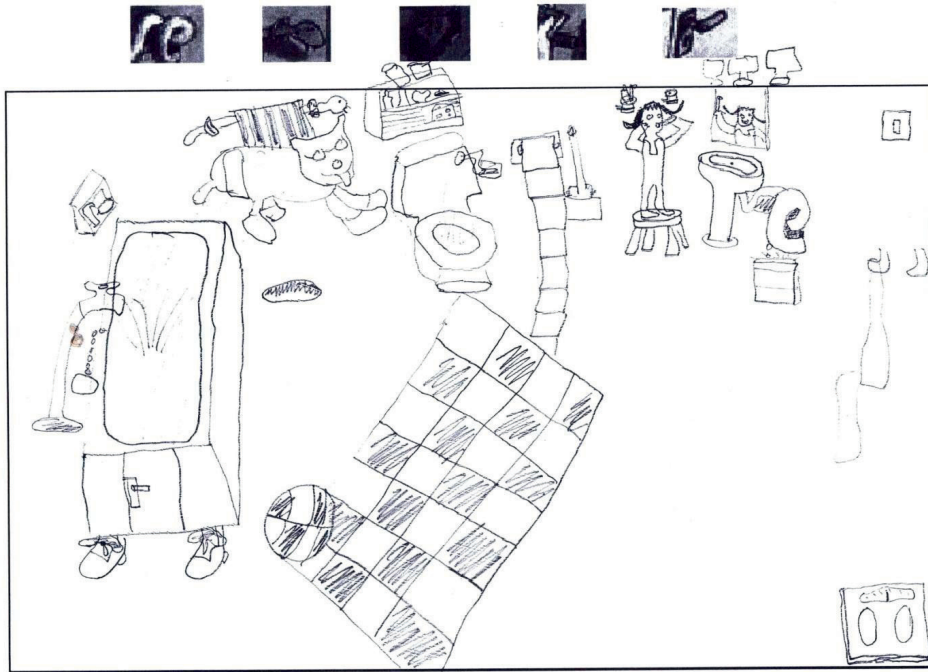
Appendix A

This appendix shows one of the illustrations used for the picture replication task (reproduced with the permission of the publisher Baumhaus Verlag), an empty sheet for drawing and two examples of the resulting drawings.









Appendix B

The following table lists the means and standard deviations of the variables observed in the three studied English function words (in Chapter 3) for groups of items based on the speakers' L1 and speaking style (CZ=Czech, EN=English, NO=Norwegian; R=read, S=spontaneous).

IN							
		CZ-R	CZ-S	EN-R	EN-S	NO-R	NO-S
Normalised duration (ms)	mean	117.8	122.0	114.0	124.2	117.7	121.2
	sd	30.6	30.5	32.5	23.2	26.1	26.4
Vowel proportion (%)	mean	43.7	48.3	44.2	43.3	46.5	46.5
	sd	11.9	13.9	12.2	13.3	13.7	11.4
Vowel nasalisation degree	mean	1.91	1.79	1.90	1.92	1.95	2.00
	sd	0.58	0.50	0.61	0.67	0.67	0.58
Following segment's nasalisation degree	mean	1.74	1.71	2.22	2.19	1.83	2.31
	sd	0.68	0.61	0.68	0.68	0.66	0.62
OF							
		CZ-R	CZ-S	EN-R	EN-S	NO-R	NO-S
Normalised duration (ms)	mean	138.1	108.6	95.7	97.4	111.1	94.7
	sd	46.3	21.9	25.2	35.3	40.2	35.7
Vowel proportion (%)	mean	38.5	52.8	42.2	51.5	55.2	53.0
	sd	12.3	13.9	16.7	22.2	15.6	18.0
Fricative voicing proportion (%)	mean	55.6	90.1	61.9	68.3	79.6	90.2
	sd	37.4	22.8	38.5	31.1	29.8	24.1
F1-F0 (Bark)	mean	3.04	2.89	2.85	2.80	3.21	2.93
	sd	0.65	0.61	0.65	0.72	0.75	0.61
F3-F2 (Bark)	mean	4.74	4.49	3.60	3.70	4.38	4.57
	sd	0.79	1.10	0.85	0.90	1.02	1.00
Friction intensity in fricative (dB)	mean	-17.9	-30.8	-17.5	-20.6	-32.2	-35.3
	sd	17.6	11.8	12.1	11.8	15.7	15.1

TO							
		CZ-R	CZ-S	EN-R	EN-S	NO-R	NO-S
Normalised duration (ms)	mean	136.8	135.4	123.1	124.0	138.0	136.6
	sd	33.7	39.6	23.2	28.9	35.6	39.4
Vowel proportion (%)	mean	31.6	29.2	19.4	22.3	27.5	28.6
	sd	7.8	13.9	12.4	12.3	8.6	12.4
Plosive release proportion (%)	mean	46.9	53.6	57.5	62.7	46.2	56.4
	sd	13.5	14.9	14.0	14.8	11.6	11.5
F1-F0 (Bark)	mean	1.96	2.02	2.32	2.19	1.99	2.15
	sd	0.48	0.50	1.11	0.87	0.58	0.60
F3-F2 (Bark)	mean	3.07	3.22	2.98	3.33	3.02	3.29
	sd	0.89	0.68	0.78	0.73	0.76	0.77
Friction intensity in plosive release (dB)	mean	-7.7	-4.1	-5.5	-3.7	-11.1	-10.7
	sd	5.8	6.7	7.2	7.1	5.6	6.7

Appendix C

The following table lists the means and standard deviations of the variables observed in repeated mentions of content words (in Chapter 4) for groups of items based on the spoken language and speakers' L1 (CZE=Czech, ENG-C=English spoken by Czech speakers, ENG-E =English spoken by native speakers, ENG-N=English spoken by Norwegian speakers, NOR=Norwegian).

Whole word duration				
		1 st mention	2 nd mention	3 rd mention
CZE	mean	478.6	427.7	434.2
	sd	199.8	180.4	163.9
ENG-C	mean	548.0	443.0	427.4
	sd	166.4	131.7	115.0
ENG-E	mean	480.0	424.6	425.1
	sd	171.2	130.1	130.4
ENG-N	mean	471.9	428.0	420.4
	sd	149.2	145.9	136.7
NOR	mean	451.9	395.7	350.4
	sd	153.0	131.9	97.5
Mean syllable duration				
		1 st mention	2 nd mention	3 rd mention
CZE	mean	237.5	213.5	220.1
	sd	96.7	85.9	84.5
ENG-C	mean	407.0	343.9	326.7
	sd	147.9	160.6	135.5
ENG-E	mean	356.7	322.4	315.6
	sd	154.6	141.2	123.4
ENG-N	mean	376.8	335.7	330.6
	sd	173.1	144.8	146.4
NOR	mean	239.3	208.7	178.3
	sd	118.8	100.8	53.6

Stressed syllable duration - monosyllabic words				
		1 st mention	2 nd mention	3 rd mention
CZE	mean	294.2	259.0	282.0
	sd	141.1	117.3	82.4
ENG-C	mean	492.1	444.8	407.2
	sd	133.2	133.5	116.2
ENG-E	mean	420.6	399.5	374.0
	sd	156.9	126.3	120.2
ENG-N	mean	456.5	401.6	394.7
	sd	152.8	129.9	135.1
NOR	mean	442.5	367.1	250.9
	sd	138.3	134.8	49.3
Stressed syllable duration - polysyllabic words				
		1 st mention	2 nd mention	3 rd mention
CZE	mean	195.2	170.7	177.8
	sd	78.4	70.4	70.0
ENG-C	mean	228.4	179.4	198.3
	sd	79.4	64.7	72.4
ENG-E	mean	238.6	197.8	211.3
	sd	73.8	47.0	57.6
ENG-N	mean	243.3	234.1	219.3
	sd	110.0	85.2	69.3
NOR	mean	214.9	184.2	173.7
	sd	78.9	74.8	49.1
Untressed syllable duration - polysyllabic words				
		1 st mention	2 nd mention	3 rd mention
CZE	mean	240.2	225.5	223.6
	sd	99.3	98.2	118.4
ENG-C	mean	334.5	221.0	225.8
	sd	117.7	96.1	89.5
ENG-E	mean	281.3	212.5	241.8
	sd	140.0	106.2	114.4
ENG-N	mean	196.9	182.0	192.6
	sd	105.3	91.9	122.0
NOR	mean	193.2	179.5	158.9
	sd	87.6	67.8	68.0

Unstressed-to-stressed syllable duration ratio				
		1 st mention	2 nd mention	3 rd mention
CZE	mean	1.34	1.45	1.41
	sd	0.66	0.70	0.87
ENG-C	mean	1.78	1.43	1.32
	sd	1.19	0.86	0.70
ENG-E	mean	1.21	1.13	1.25
	sd	0.46	0.61	0.76
ENG-N	mean	0.96	0.88	1.03
	sd	0.65	0.60	0.89
NOR	mean	0.99	1.04	0.98
	sd	0.52	0.42	0.46
Distance to the centroid				
		1 st mention	2 nd mention	3 rd mention
CZE	mean	2.15	2.01	2.02
	sd	0.80	0.61	0.69
ENG-C	mean	2.68	2.47	2.28
	sd	0.82	0.76	0.82
ENG-E	mean	2.82	2.71	2.64
	sd	0.84	0.93	0.88
ENG-N	mean	2.59	2.43	2.35
	sd	0.98	1.02	1.06
NOR	mean	2.68	2.50	2.39
	sd	1.06	1.04	0.91

Appendix D

The following table lists all lexical items occurring in the study of repeated mentions of content words (Chapter 4). The table is divided into sections based on the number of the word's syllables. Each section lists words in Czech, English and Norwegian in alphabetical order.

Monosyllabic words						
Czech	pes	plot	pouť	roh	zdi	
English	back	dot	lamp	pump	snake	track
	ball	drain	leg	roof	square	train
	bar	duck	legs	room	stall	wall
	bath	fork	line	seat	store	walls
	booth	goose	pegs	shed	tail	zed
	cans	hand	pigs	shoes	tap	
	dog	horse	plug	shop	tins	
	door	hose	poles	sink	toys	
Norwegian	bånnen	bil	hest	kant	speil	
Bisyllabic words						
Czech	dveře	rohu	svěrák	vanou	vláček	
	paní	skříňka	vana	vidle	záchod	
English	armchair	bucket	handle	piping	rhino	TV
	balloons	corner	L-shape	quarters	table	window
	beachballs	fascia	mirror	railway	toilet	woman
Norwegian	bildet	gjerde/gjerdet	hylla	ramma	TV	veggen
	boden	håndkle	hylle	taket	TVen	vogna
	gitter	hjørne	neven	teppe	vasken	vogner

Words with three or four syllables						
Czech	botičky	koryto	obličej	ohradu	prkýnka	zvířata
	holčička	krabice	ohrada	podlaha	strojvůdce	
English	animals	elephant	letterbox	merry-go-round		
	carousel	harmonica	locomotive			
Norwegian	badekar	dørhåndtak	innhegning	innhegninga	karusellen	tivoli

Appendix E

The following two sections present details of statistical analyses carried out in order to evaluate the effects of various factors, including the control factor *finality*, on the dependent variables item duration (Dur.item) and vowel distance to the centroid (Centr.dist) in the *LI data* subset. The commands entered in the statistical software R are printed in red, while the outputs from the statistical software are in black.

Item duration (Dur.item)

```
(lmer(Dur.item ~ (1|Speaker) + (1|Word), data = nat)) ->
NF_Dur.item_0a
(lmer(Dur.item ~ (1|Word), data = nat)) -> NF_Dur.item_0b
(lmer(Dur.item ~ (1|Speaker), data = nat)) -> NF_Dur.item_0c

> anova(NF_Dur.item_0a, NF_Dur.item_0b)
Data: nat
Models:
NF_Dur.item_0b: Dur.item ~ (1 | Word)
NF_Dur.item_0a: Dur.item ~ (1 | Speaker) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_0b  3 3766.8 3777.8 -1880.4
NF_Dur.item_0a  4 3767.1 3781.9 -1879.6 1.6096      1    0.2046

> anova(NF_Dur.item_0a, NF_Dur.item_0c)
Data: nat
Models:
NF_Dur.item_0c: Dur.item ~ (1 | Speaker)
NF_Dur.item_0a: Dur.item ~ (1 | Speaker) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_0c  3 3826.9 3838.0 -1910.5
NF_Dur.item_0a  4 3767.1 3781.9 -1879.6 61.76      1    3.88e-15 ***
```

```

Base model: Dur.item ~ (1 | Word)
NF_Dur.item_0b -> NF_Dur.item_0

(lmer(Dur.item ~ L1 + (1|Word), data = nat)) -> NF_Dur.item_1a
(lmer(Dur.item ~ Mention + (1|Word), data = nat)) -> NF_Dur.item_1b
(lmer(Dur.item ~ Fin + (1|Word), data = nat)) -> NF_Dur.item_1c
(lmer(Dur.item ~ log(Art.rate) + (1|Word), data = nat)) ->
NF_Dur.item_1d

> anova(NF_Dur.item_0, NF_Dur.item_1a)
Data: nat
Models:
NF_Dur.item_0: Dur.item ~ (1 | Word)
NF_Dur.item_1a: Dur.item ~ L1 + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_0  3 3766.8 3777.8 -1880.4
NF_Dur.item_1a  5 3767.2 3785.7 -1878.6  3.5469      2    0.1697

> anova(NF_Dur.item_0, NF_Dur.item_1b)
Data: nat
Models:
NF_Dur.item_0: Dur.item ~ (1 | Word)
NF_Dur.item_1b: Dur.item ~ Mention + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_0  3 3766.8 3777.8 -1880.4
NF_Dur.item_1b  5 3748.7 3767.1 -1869.3 22.086      2    1.6e-05 ***

> anova(NF_Dur.item_0, NF_Dur.item_1c)
Data: nat
Models:
NF_Dur.item_0: Dur.item ~ (1 | Word)
NF_Dur.item_1c: Dur.item ~ Fin + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_0  3 3766.8 3777.8 -1880.4
NF_Dur.item_1c  4 3731.4 3746.2 -1861.7 37.337      1    9.94e-10 ***

> anova(NF_Dur.item_0, NF_Dur.item_1d)
Data: nat
Models:
NF_Dur.item_0: Dur.item ~ (1 | Word)
NF_Dur.item_1d: Dur.item ~ log(Art.rate) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_0  3 3766.8 3777.8 -1880.4
NF_Dur.item_1d  4 3730.4 3745.2 -1861.2 38.341      1    5.939e-10 ***

Model 1: Dur.item ~ log(Art.rate) + (1 | Word)
NF_Dur.item_1d -> NF_Dur.item_1

```

```
(lmer(Dur.item ~ L1 + log(Art.rate) + (1|Word), data = nat)) ->
NF_Dur.item_2a
(lmer(Dur.item ~ Mention + log(Art.rate) + (1|Word), data = nat)) ->
NF_Dur.item_2b
(lmer(Dur.item ~ Fin + log(Art.rate) + (1|Word), data = nat)) ->
NF_Dur.item_2c
```

```
> anova(NF_Dur.item_1, NF_Dur.item_2a)
Data: nat
Models:
NF_Dur.item_1: Dur.item ~ log(Art.rate) + (1 | Word)
NF_Dur.item_2a: Dur.item ~ L1 + log(Art.rate) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_1  4 3730.4 3745.2 -1861.2
NF_Dur.item_2a 6 3731.3 3753.4 -1859.6 3.1455      2 0.2075
```

```
> anova(NF_Dur.item_1, NF_Dur.item_2b)
Data: nat
Models:
NF_Dur.item_1: Dur.item ~ log(Art.rate) + (1 | Word)
NF_Dur.item_2b: Dur.item ~ Mention + log(Art.rate) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_1  4 3730.4 3745.2 -1861.2
NF_Dur.item_2b 6 3711.4 3733.5 -1849.7 23.049      2 9.885e-06 ***
```

```
> anova(NF_Dur.item_1, NF_Dur.item_2c)
Data: nat
Models:
NF_Dur.item_1: Dur.item ~ log(Art.rate) + (1 | Word)
NF_Dur.item_2c: Dur.item ~ Fin + log(Art.rate) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_1  4 3730.4 3745.2 -1861.2
NF_Dur.item_2c 5 3698.9 3717.3 -1844.4 33.541      1 6.976e-09 ***
```

Model 2: Dur.item ~ Fin + log(Art.rate) + (1 | Word)

```
NF_Dur.item_2c -> NF_Dur.item_2
```

```
(lmer(Dur.item ~ L1 + Fin + log(Art.rate) + (1|Word), data = nat)) ->
NF_Dur.item_3a
(lmer(Dur.item ~ Mention + Fin + log(Art.rate) + (1|Word), data =
nat)) -> NF_Dur.item_3b
(lmer(Dur.item ~ Fin * log(Art.rate) + (1|Word), data = nat)) ->
NF_Dur.item_3c
```

```
> anova(NF_Dur.item_2, NF_Dur.item_3a)
Data: nat
Models:
NF_Dur.item_2: Dur.item ~ Fin + log(Art.rate) + (1 | Word)
NF_Dur.item_3a: Dur.item ~ L1 + Fin + log(Art.rate) + (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Dur.item_2	5	3698.9	3717.3	-1844.4				
NF_Dur.item_3a	7	3698.8	3724.6	-1842.4	4.0651		2	0.131

```
> anova(NF_Dur.item_2, NF_Dur.item_3b)
```

```
Data: nat
```

```
Models:
```

```
NF_Dur.item_2: Dur.item ~ Fin + log(Art.rate) + (1 | Word)
```

```
NF_Dur.item_3b: Dur.item ~ Mention + Fin + log(Art.rate) + (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Dur.item_2	5	3698.9	3717.3	-1844.4				
NF_Dur.item_3b	7	3683.5	3709.3	-1834.7	19.412		2	6.093e-05 ***

```
> anova(NF_Dur.item_2, NF_Dur.item_3c)
```

```
Data: nat
```

```
Models:
```

```
NF_Dur.item_2: Dur.item ~ Fin + log(Art.rate) + (1 | Word)
```

```
NF_Dur.item_3c: Dur.item ~ Fin * log(Art.rate) + (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Dur.item_2	5	3698.9	3717.3	-1844.4				
NF_Dur.item_3c	6	3692.9	3715.0	-1840.4	7.9821		1	0.004724 **

```
Model 3: Dur.item ~ Dur.item ~ Mention + Fin + log(Art.rate) + (1 | Word)
```

```
NF_Dur.item_3b -> NF_Dur.item_3
```

```
(lmer(Dur.item ~ L1 + Mention + Fin + log(Art.rate) + (1|Word), data = nat)) -> NF_Dur.item_4a
```

```
(lmer(Dur.item ~ Mention * Fin + log(Art.rate) + (1|Word), data = nat)) -> NF_Dur.item_4b
```

```
(lmer(Dur.item ~ Mention + Fin * log(Art.rate) + (1|Word), data = nat)) -> NF_Dur.item_4c
```

```
(lmer(Dur.item ~ Mention * log(Art.rate) + Fin + (1|Word), data = nat)) -> NF_Dur.item_4d
```

```
> anova(NF_Dur.item_3, NF_Dur.item_4a)
```

```
Data: nat
```

```
Models:
```

```
NF_Dur.item_3: Dur.item ~ Mention + Fin + log(Art.rate) + (1 | Word)
```

```
NF_Dur.item_4a: Dur.item ~ L1 + Mention + Fin + log(Art.rate) + (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Dur.item_3	7	3683.5	3709.3	-1834.7				
NF_Dur.item_4a	9	3683.3	3716.5	-1832.7	4.139		2	0.1262

```
> anova(NF_Dur.item_3, NF_Dur.item_4b)
```

```
Data: nat
```

```
Models:
```

```
NF_Dur.item_3: Dur.item ~ Mention + Fin + log(Art.rate) + (1 | Word)
```

```

NF_Dur.item_4b: Dur.item ~ Mention * Fin + log(Art.rate) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_3  7 3683.5 3709.3 -1834.7
NF_Dur.item_4b  9 3685.8 3719.0 -1833.9 1.6086      2    0.4474

```

```
> anova(NF_Dur.item_3, NF_Dur.item_4c)
```

```

Data: nat
Models:
NF_Dur.item_3: Dur.item ~ Mention + Fin + log(Art.rate) + (1 | Word)
NF_Dur.item_4c: Dur.item ~ Mention + Fin * log(Art.rate) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_3  7 3683.5 3709.3 -1834.7
NF_Dur.item_4c  8 3676.8 3706.3 -1830.4 8.6636      1    0.003246 **

```

```
> anova(NF_Dur.item_3, NF_Dur.item_4d)
```

```

Data: nat
Models:
NF_Dur.item_3: Dur.item ~ Mention + Fin + log(Art.rate) + (1 | Word)
NF_Dur.item_4d: Dur.item ~ Mention * log(Art.rate) + Fin + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_3  7 3683.5 3709.3 -1834.7
NF_Dur.item_4d  9 3687.4 3720.6 -1834.7 0.0604      2    0.9703

```

```
Model 4: Dur.item ~ Dur.item ~ Mention + Fin * log(Art.rate) + (1 | Word)
```

```
NF_Dur.item_4c -> NF_Dur.item_4
```

```

(lmer(Dur.item ~ L1 + Mention + Fin * log(Art.rate) + (1|Word), data =
nat)) -> NF_Dur.item_5a
(lmer(Dur.item ~ Mention * Fin * log(Art.rate) + (1|Word), data =
nat)) -> NF_Dur.item_5b
(lmer(Dur.item ~ Mention + Fin * log(Art.rate) + (Mention:Fin) +
(1|Word), data = nat)) -> NF_Dur.item_5c
(lmer(Dur.item ~ Mention + Fin * log(Art.rate) +
(Mention:log(Art.rate)) + (1|Word), data = nat)) -> NF_Dur.item_5d

```

```
> anova(NF_Dur.item_4, NF_Dur.item_5a)
```

```

Data: nat
Models:
NF_Dur.item_4: Dur.item ~ Mention + Fin * log(Art.rate) + (1 | Word)
NF_Dur.item_5a: Dur.item ~ L1 + Mention + Fin * log(Art.rate) + (1 |
Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
NF_Dur.item_4  8 3676.8 3706.3 -1830.4
NF_Dur.item_5a 10 3676.9 3713.7 -1828.4 3.9297      2    0.1402

```

```
> anova(NF_Dur.item_4, NF_Dur.item_5b)
```

```

Data: nat
Models:

```

```

NF_Dur.item_4: Dur.item ~ Mention + Fin * log(Art.rate) + (1 | Word)
NF_Dur.item_5b: Dur.item ~ Mention * Fin * log(Art.rate) + (1 | Word)
      Df    AIC    BIC logLik Chisq Chi Df Pr(>Chisq)
NF_Dur.item_4   8 3676.8 3706.3 -1830.4
NF_Dur.item_5b 14 3683.8 3735.4 -1827.9 4.968      6    0.5479

```

```
> anova(NF_Dur.item_4, NF_Dur.item_5c)
```

```
Data: nat
```

```
Models:
```

```

NF_Dur.item_4: Dur.item ~ Mention + Fin * log(Art.rate) + (1 | Word)
NF_Dur.item_5c: Dur.item ~ Mention + Fin * log(Art.rate) +
(Mention:Fin) + (1 | Word)
      Df    AIC    BIC logLik Chisq Chi Df Pr(>Chisq)
NF_Dur.item_4   8 3676.8 3706.3 -1830.4
NF_Dur.item_5c 10 3679.1 3716.0 -1829.5 1.6897      2    0.4296

```

```
> anova(NF_Dur.item_4, NF_Dur.item_5d)
```

```
Data: nat
```

```
Models:
```

```

NF_Dur.item_4: Dur.item ~ Mention + Fin * log(Art.rate) + (1 | Word)
NF_Dur.item_5d: Dur.item ~ Mention + Fin * log(Art.rate) +
(Mention:log(Art.rate)) +
NF_Dur.item_5d: (1 | Word)
      Df    AIC    BIC logLik Chisq Chi Df Pr(>Chisq)
NF_Dur.item_4   8 3676.8 3706.3 -1830.4
NF_Dur.item_5d 10 3680.3 3717.2 -1830.2 0.4697      2    0.7907

```

Model 4: Dur.item ~ Mention + Fin * log(Art.rate) + (1 | Word)

```
> summary(NF_Dur.item_4)
```

```
Linear mixed model fit by REML
```

```
Formula: Dur.item ~ Mention + Fin * log(Art.rate) + (1 | Word)
```

```
Data: nat
```

```

      AIC    BIC logLik deviance REMLdev
3632 3661 -1808    3661    3616

```

```
Random effects:
```

```

Groups   Name          Variance Std.Dev.
Word    (Intercept) 12688.1 112.642
Residual                8878.8  94.227

```

```
Number of obs: 295, groups: Word, 84
```

```
Fixed effects:
```

```

              Estimate Std. Error t value
(Intercept)    551.69    47.73  11.559
Mention2        -44.15    13.65  -3.235
Mention3        -59.33    13.60  -4.362
Finnf           137.37    73.51   1.869
log(Art.rate)  -23.38    26.33  -0.888
Finnf:log(Art.rate) -121.91    41.49  -2.939

```

```

Correlation of Fixed Effects:
      (Intr) Mention2  Mention3  Finnf  lg(A.)
Mention2      -0.221
Mention3      -0.090  0.490
Finnf         -0.584  0.121 -0.091
log(Art.rt)   -0.934  0.095 -0.032  0.584
Fnnf:lg(A.)   0.588 -0.135  0.068 -0.984 -0.615

```

```
> anova(NF_Dur.item_4)
```

```
Analysis of Variance Table
```

	Df	Sum Sq	Mean Sq	F value
Mention	2	253499	126749	14.2756
Fin	1	348597	348597	39.2619
log(Art.rate)	1	103758	103758	11.6861
Fin:log(Art.rate)	1	76673	76673	8.6356

Distance to the centroid (Centr.dist)

```

(lmer(Centr.dist ~ (1|Speaker) + (1|Word), data = nat.vowels)) ->
NF_Centr.dist_0a
(lmer(Centr.dist ~ (1|Word), data = nat.vowels)) -> NF_Centr.dist_0b
(lmer(Centr.dist ~ (1|Speaker), data = nat.vowels)) ->
NF_Centr.dist_0c

```

```
> anova(NF_Centr.dist_0a, NF_Centr.dist_0b)
```

```
Data: nat.vowels
```

```
Models:
```

```
NF_Centr.dist_0b: Centr.dist ~ (1 | Word)
```

```
NF_Centr.dist_0a: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_0b	3	382.70	392.57	-188.35				
NF_Centr.dist_0a	4	375.59	388.74	-183.79	9.1167		1	0.002533 **

```
> anova(NF_Centr.dist_0a, NF_Centr.dist_0c)
```

```
Data: nat.vowels
```

```
Models:
```

```
NF_Centr.dist_0c: Centr.dist ~ (1 | Speaker)
```

```
NF_Centr.dist_0a: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_0c	3	495.37	505.24	-244.69				
NF_Centr.dist_0a	4	375.59	388.74	-183.79	121.78		1	< 2.2e-16 ***

```
Base model: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
NF_Centr.dist_0a -> NF_Centr.dist_0
```



```
(lmer(Centr.dist ~ L1 + (1 | Speaker) + (1|Word), data = nat.vowels))
-> NF_Centr.dist_1a
(lmer(Centr.dist ~ Mention + (1 | Speaker) + (1|Word), data =
nat.vowels)) -> NF_Centr.dist_1b
(lmer(Centr.dist ~ Fin + (1 | Speaker) + (1|Word), data = nat.vowels))
-> NF_Centr.dist_1c
(lmer(Centr.dist ~ log(Art.rate) + (1 | Speaker) + (1|Word), data =
nat.vowels)) -> NF_Centr.dist_1d
```

```
> anova(NF_Centr.dist_0, NF_Centr.dist_1a)
```

```
Data: nat.vowels
```

```
Models:
```

```
NF_Centr.dist_0: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
NF_Centr.dist_1a: Centr.dist ~ L1 + (1 | Speaker) + (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_0	4	375.59	388.74	-183.79				
NF_Centr.dist_1a	6	375.52	395.25	-181.76	4.0711		2	0.1306

```
> anova(NF_Centr.dist_0, NF_Centr.dist_1b)
```

```
Data: nat.vowels
```

```
Models:
```

```
NF_Centr.dist_0: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
NF_Centr.dist_1b: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_0	4	375.59	388.74	-183.79				
NF_Centr.dist_1b	6	370.82	390.55	-179.41	8.7657		2	0.01249 *

```
> anova(NF_Centr.dist_0, NF_Centr.dist_1c)
```

```
Data: nat.vowels
```

```
Models:
```

```
NF_Centr.dist_0: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
NF_Centr.dist_1c: Centr.dist ~ Fin + (1 | Speaker) + (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_0	4	375.59	388.74	-183.79				
NF_Centr.dist_1c	5	377.36	393.81	-183.68	0.2238		1	0.6362

```
> anova(NF_Centr.dist_0, NF_Centr.dist_1d)
```

```
Data: nat.vowels
```

```
Models:
```

```
NF_Centr.dist_0: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
NF_Centr.dist_1d: Centr.dist ~ log(Art.rate) + (1 | Speaker) + (1 |
Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_0	4	375.59	388.74	-183.79				
NF_Centr.dist_1d	5	374.57	390.99	-182.29	3.015		1	0.0825 .

```
Model 1: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)
```

```
NF_Centr.dist_1b -> NF_Centr.dist_1
```

```
(lmer(Centr.dist ~ L1 + Mention + (1 | Speaker) + (1|Word), data =
nat.vowels)) -> NF_Centr.dist_2a
(lmer(Centr.dist ~ Fin + Mention + (1 | Speaker) + (1|Word), data =
nat.vowels)) -> NF_Centr.dist_2b
(lmer(Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker) + (1|Word),
data = nat.vowels)) -> NF_Centr.dist_2c
```

```
> anova(NF_Centr.dist_1, NF_Centr.dist_2a)
```

```
Data: nat.vowels
```

```
Models:
```

```
NF_Centr.dist_1: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)
```

```
NF_Centr.dist_2a: Centr.dist ~ L1 + Mention + (1 | Speaker) + (1 |
Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_1	6	370.82	390.55	-179.41				
NF_Centr.dist_2a	8	370.77	397.08	-177.38	4.0524		2	0.1318

```
> anova(NF_Centr.dist_1, NF_Centr.dist_2b)
```

```
Data: nat.vowels
```

```
Models:
```

```
NF_Centr.dist_1: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)
```

```
NF_Centr.dist_2b: Centr.dist ~ Fin + Mention + (1 | Speaker) + (1 |
Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_1	6	370.82	390.55	-179.41				
NF_Centr.dist_2b	7	372.11	395.13	-179.06	0.7102		1	0.3994

```
> anova(NF_Centr.dist_1, NF_Centr.dist_2c)
```

```
Data: nat.vowels
```

```
Models:
```

```
NF_Centr.dist_1: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)
```

```
NF_Centr.dist_2c: Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker)
+ (1 | Word)
```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_1	6	370.82	390.55	-179.41				
NF_Centr.dist_2c	7	368.26	391.24	-177.13	4.5644		1	0.03264 *

```
Model 2: Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker) + (1 |
Word)
```

```
NF_Centr.dist_2c -> NF_Centr.dist_2
```

```
(lmer(Centr.dist ~ L1 + log(Art.rate) + Mention + (1 | Speaker) +
(1|Word), data = nat.vowels)) -> NF_Centr.dist_3a
```

```
(lmer(Centr.dist ~ Fin + log(Art.rate) + Mention + (1 | Speaker) +
(1|Word), data = nat.vowels)) -> NF_Centr.dist_3b
```

```
(lmer(Centr.dist ~ log(Art.rate) * Mention + (1 | Speaker) + (1|Word),
data = nat.vowels)) -> NF_Centr.dist_3c
```

```

> anova(NF_Centr.dist_2, NF_Centr.dist_3a)
Data: nat.vowels
Models:
NF_Centr.dist_2: Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker)
+ (1 | Word)
NF_Centr.dist_3a: Centr.dist ~ L1 + log(Art.rate) + Mention + (1 |
Speaker) + (1 | Word)

```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_2	7	368.26	391.24	-177.13				
NF_Centr.dist_3a	9	368.78	398.33	-175.39	3.4731		2	0.1761

```

> anova(NF_Centr.dist_2, NF_Centr.dist_3b)
Data: nat.vowels
Models:
NF_Centr.dist_2: Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker)
+ (1 | Word)
NF_Centr.dist_3b: Centr.dist ~ Fin + log(Art.rate) + Mention + (1 |
Speaker) + (1 | Word)

```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_2	7	368.26	391.24	-177.13				
NF_Centr.dist_3b	8	369.52	395.78	-176.76	0.7424		1	0.3889

```

> anova(NF_Centr.dist_2, NF_Centr.dist_3c)
Data: nat.vowels
Models:
NF_Centr.dist_2: Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker)
+ (1 | Word)
NF_Centr.dist_3c: Centr.dist ~ log(Art.rate) * Mention + (1 | Speaker)
+ (1 | Word)

```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
NF_Centr.dist_2	7	368.26	391.24	-177.13				
NF_Centr.dist_3c	9	371.72	401.27	-176.86	0.5404		2	0.7632

Model 2: Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker) + (1 | Word)

```

> summary(NF_Centr.dist_2)
Linear mixed model fit by REML
Formula: Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker) + (1 |
Word)
Data: nat.vowels
AIC BIC logLik deviance REMLdev
380.5 403.5 -183.2 354.3 366.5
Random effects:
Groups Name Variance Std.Dev.
Word (Intercept) 0.59553 0.77170
Speaker (Intercept) 0.13808 0.37159
Residual 0.15348 0.39177
Number of obs: 197, groups: Word, 57; Speaker, 26

```

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.78519	0.23288	11.960
log(Art.rate)	-0.14857	0.10927	-1.360
Mention2	-0.15376	0.06867	-2.239
Mention3	-0.21664	0.06867	-3.155

Correlation of Fixed Effects:

	(Intr)	lg(A.)	Mention2
log(Art.rt)	-0.811		
Mention2	-0.190	0.051	
Mention3	-0.192	0.053	0.507

> anova(NF_Centr.dist_2)

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value
log(Art.rate)	1	0.21152	0.21152	1.3781
Mention	2	1.61207	0.80604	5.2516

Appendix F

The following two sections present details of statistical analyses carried out in order to evaluate the effects of various factors, including the control factor *finality*, on the dependent variables item duration (Dur.item) and vowel distance to the centroid (Centr.dist) in the *English data* subset. The commands entered in the statistical software R are printed in red, while the outputs from the statistical software are in black.

Item duration (Dur.item)

```
(lmer(Dur.item ~ (1|Speaker) + (1|Word), data = eng)) ->
EF_Dur.item_0a
(lmer(Dur.item ~ (1|Word), data = eng)) -> EF_Dur.item_0b
(lmer(Dur.item ~ (1|Speaker), data = eng)) -> EF_Dur.item_0c

> anova(EF_Dur.item_0a, EF_Dur.item_0b)
Data: eng
Models:
EF_Dur.item_0b: Dur.item ~ (1 | Word)
EF_Dur.item_0a: Dur.item ~ (1 | Speaker) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Dur.item_0b  3 3878.3 3889.5 -1936.2
EF_Dur.item_0a  4 3857.8 3872.7 -1924.9 22.487      1 2.115e-06 ***

> anova(EF_Dur.item_0a, EF_Dur.item_0c)
Data: eng
Models:
EF_Dur.item_0c: Dur.item ~ (1 | Speaker)
EF_Dur.item_0a: Dur.item ~ (1 | Speaker) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Dur.item_0c  3 3883.0 3894.2 -1938.5
EF_Dur.item_0a  4 3857.8 3872.7 -1924.9 27.204      1 1.831e-07 ***
```

Base model: Dur.item ~ (1 | Speaker) + (1 | Word)

EF_Dur.item_0a -> EF_Dur.item_0

```
(lmer(Dur.item ~ L1 + (1 | Speaker) + (1|Word), data = eng)) ->
EF_Dur.item_1a
(lmer(Dur.item ~ Mention + (1 | Speaker) + (1|Word), data = eng)) ->
EF_Dur.item_1b
(lmer(Dur.item ~ Fin + (1 | Speaker) + (1|Word), data = eng)) ->
EF_Dur.item_1c
(lmer(Dur.item ~ log(Art.rate) + (1 | Speaker) + (1|Word), data =
eng)) -> EF_Dur.item_1d
```

> anova(EF_Dur.item_0, EF_Dur.item_1a)

Data: eng

Models:

EF_Dur.item_0: Dur.item ~ (1 | Speaker) + (1 | Word)

EF_Dur.item_1a: Dur.item ~ L1 + (1 | Speaker) + (1 | Word)

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
EF_Dur.item_0	4	3857.8	3872.7	-1924.9				
EF_Dur.item_1a	6	3860.8	3883.2	-1924.4	0.9797		2	0.6127

> anova(EF_Dur.item_0, EF_Dur.item_1b)

Data: eng

Models:

EF_Dur.item_0: Dur.item ~ (1 | Speaker) + (1 | Word)

EF_Dur.item_1b: Dur.item ~ Mention + (1 | Speaker) + (1 | Word)

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
EF_Dur.item_0	4	3857.8	3872.7	-1924.9				
EF_Dur.item_1b	6	3831.6	3853.9	-1909.8	30.209		2	2.756e-07 ***

> anova(EF_Dur.item_0, EF_Dur.item_1c)

Data: eng

Models:

EF_Dur.item_0: Dur.item ~ (1 | Speaker) + (1 | Word)

EF_Dur.item_1c: Dur.item ~ Fin + (1 | Speaker) + (1 | Word)

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
EF_Dur.item_0	4	3857.8	3872.7	-1924.9				
EF_Dur.item_1c	5	3801.6	3820.2	-1895.8	58.251		1	2.308e-14 ***

> anova(EF_Dur.item_0, EF_Dur.item_1d)

Data: eng

Models:

EF_Dur.item_0: Dur.item ~ (1 | Speaker) + (1 | Word)

EF_Dur.item_1d: Dur.item ~ log(Art.rate) + (1 | Speaker) + (1 | Word)

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
EF_Dur.item_0	4	3857.8	3872.7	-1924.9				
EF_Dur.item_1d	5	3795.4	3814.0	-1892.7	64.396		1	1.017e-15 ***

Model 1: Dur.item ~ log(Art.rate) + (1 | Speaker) + (1 | Word)

```

EF_Dur.item_1d -> EF_Dur.item_1

(lmer(Dur.item ~ L1 + log(Art.rate) + (1 | Speaker) + (1|Word), data =
eng)) -> EF_Dur.item_2a
(lmer(Dur.item ~ Mention + log(Art.rate) + (1 | Speaker) + (1|Word),
data = eng)) -> EF_Dur.item_2b
(lmer(Dur.item ~ Fin + log(Art.rate) + (1 | Speaker) + (1|Word), data
= eng)) -> EF_Dur.item_2c

> anova(EF_Dur.item_1, EF_Dur.item_2a)
Data: eng
Models:
EF_Dur.item_1: Dur.item ~ log(Art.rate) + (1 | Speaker) + (1 | Word)
EF_Dur.item_2a: Dur.item ~ L1 + log(Art.rate) + (1 | Speaker) + (1 |
Word)
          Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Dur.item_1    5 3795.4 3814.0 -1892.7
EF_Dur.item_2a   7 3798.7 3824.6 -1892.3 0.7417    2    0.6901

> anova(EF_Dur.item_1, EF_Dur.item_2b)
Data: eng
Models:
EF_Dur.item_1: Dur.item ~ log(Art.rate) + (1 | Speaker) + (1 | Word)
EF_Dur.item_2b: Dur.item ~ Mention + log(Art.rate) + (1 | Speaker) +
(1 | Word)
          Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Dur.item_1    5 3795.4 3814.0 -1892.7
EF_Dur.item_2b   7 3771.2 3797.2 -1878.6 28.233    2 7.401e-07 ***

> anova(EF_Dur.item_1, EF_Dur.item_2c)
Data: eng
Models:
EF_Dur.item_1: Dur.item ~ log(Art.rate) + (1 | Speaker) + (1 | Word)
EF_Dur.item_2c: Dur.item ~ Fin + log(Art.rate) + (1 | Speaker) + (1 |
Word)
          Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Dur.item_1    5 3795.4 3814  -1892.7
EF_Dur.item_2c   6 3745.8 3768 -1866.9 51.66    1 6.598e-13 ***

Model 2: Dur.item ~ Fin + log(Art.rate) + (1 | Speaker) + (1 | Word)
EF_Dur.item_2c -> EF_Dur.item_2

(lmer(Dur.item ~ L1 + Fin + log(Art.rate) + (1 | Speaker) + (1|Word),
data = eng)) -> EF_Dur.item_3a
(lmer(Dur.item ~ Mention + Fin + log(Art.rate) + (1 | Speaker) +
(1|Word), data = eng)) -> EF_Dur.item_3b
(lmer(Dur.item ~ Fin * log(Art.rate) + (1 | Speaker) + (1|Word), data
= eng)) -> EF_Dur.item_3c

```

```

> anova(EF_Dur.item_2, EF_Dur.item_3a)
Data: eng
Models:
EF_Dur.item_2: Dur.item ~ Fin + log(Art.rate) + (1 | Speaker) + (1 |
Word)
EF_Dur.item_3a: Dur.item ~ L1 + Fin + log(Art.rate) + (1 | Speaker) +
(1 | Word)

```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
EF_Dur.item_2	6	3745.8	3768.0	-1866.9				
EF_Dur.item_3a	8	3749.2	3778.8	-1866.6	0.5926		2	0.7435

```

> anova(EF_Dur.item_2, EF_Dur.item_3b)
Data: eng
Models:
EF_Dur.item_2: Dur.item ~ Fin + log(Art.rate) + (1 | Speaker) + (1 |
Word)
EF_Dur.item_3b: Dur.item ~ Mention + Fin + log(Art.rate) + (1 |
Speaker) + (1 | Word)

```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
EF_Dur.item_2	6	3745.8	3768	-1866.9				
EF_Dur.item_3b	8	3727.3	3757	-1855.7	22.421		2	1.353e-05 ***

```

> anova(EF_Dur.item_2, EF_Dur.item_3c)
Data: eng
Models:
EF_Dur.item_2: Dur.item ~ Fin + log(Art.rate) + (1 | Speaker) + (1 |
Word)
EF_Dur.item_3c: Dur.item ~ Fin * log(Art.rate) + (1 | Speaker) + (1 |
Word)

```

	Df	AIC	BIC	logLik	Chisq	Chi	Df	Pr(>Chisq)
EF_Dur.item_2	6	3745.8	3768.0	-1866.9				
EF_Dur.item_3c	7	3747.4	3773.4	-1866.7	0.3292		1	0.5661

```

Model 3: Dur.item ~ Dur.item ~ Mention + Fin + log(Art.rate) +
(1|speaker) + (1 | Word)
EF_Dur.item_3b -> EF_Dur.item_3

```

```

(lmer(Dur.item ~ L1 + Mention + Fin + log(Art.rate) + (1 | Speaker) +
(1|Word), data = eng)) -> EF_Dur.item_4a
(lmer(Dur.item ~ Mention * Fin + log(Art.rate) + (1 | Speaker) +
(1|Word), data = eng)) -> EF_Dur.item_4b
(lmer(Dur.item ~ Mention + Fin * log(Art.rate) + (1 | Speaker) +
(1|Word), data = eng)) -> EF_Dur.item_4c
(lmer(Dur.item ~ Mention * log(Art.rate) + Fin + (1 | Speaker) +
(1|Word), data = eng)) -> EF_Dur.item_4d

```

```

> anova(EF_Dur.item_3, EF_Dur.item_4a)

```



```
Data: eng
Models:
EF_Dur.item_3: Dur.item ~ Mention + Fin + log(Art.rate) + (1 |
Speaker) + (1 | Word)
EF_Dur.item_4a: Dur.item ~ L1 + Mention + Fin + log(Art.rate) + (1 |
Speaker) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Dur.item_3    8 3727.3 3757.0 -1855.7
EF_Dur.item_4a  10 3730.7 3767.8 -1855.3 0.6811    2    0.7114
```

```
> anova(EF_Dur.item_3, EF_Dur.item_4b)
```

```
Data: eng
Models:
EF_Dur.item_3: Dur.item ~ Mention + Fin + log(Art.rate) + (1 |
Speaker) + (1 | Word)
EF_Dur.item_4b: Dur.item ~ Mention * Fin + log(Art.rate) + (1 |
Speaker) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Dur.item_3    8 3727.3 3757.0 -1855.7
EF_Dur.item_4b  10 3729.4 3766.5 -1854.7 1.9647    2    0.3744
```

```
> anova(EF_Dur.item_3, EF_Dur.item_4c)
```

```
Data: eng
Models:
EF_Dur.item_3: Dur.item ~ Mention + Fin + log(Art.rate) + (1 |
Speaker) + (1 | Word)
EF_Dur.item_4c: Dur.item ~ Mention + Fin * log(Art.rate) + (1 |
Speaker) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Dur.item_3    8 3727.3 3757.0 -1855.7
EF_Dur.item_4c    9 3728.8 3762.2 -1855.4 0.5714    1    0.4497
```

```
> anova(EF_Dur.item_3, EF_Dur.item_4d)
```

```
Data: eng
Models:
EF_Dur.item_3: Dur.item ~ Mention + Fin + log(Art.rate) + (1 |
Speaker) + (1 | Word)
EF_Dur.item_4d: Dur.item ~ Mention * log(Art.rate) + Fin + (1 |
Speaker) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Dur.item_3    8 3727.3 3757 -1855.7
EF_Dur.item_4d  10 3730.9 3768 -1855.4 0.4496    2    0.7987
```

```
Model 3: Dur.item ~ Mention + Fin + log(Art.rate) + (1 | Speaker) + (1 | Word)
```

```
> summary(EF_Dur.item_3)
```

```
Linear mixed model fit by REML
```

```

Formula: Dur.item ~ Mention + Fin + log(Art.rate) + (1 | Speaker) + (1
| Word)
Data: eng
  AIC BIC logLik deviance REMLdev
3692 3722 -1838    3711    3676
Random effects:
Groups   Name             Variance Std.Dev.
Word     (Intercept)  5061.4   71.144
Speaker  (Intercept)  3529.0   59.406
Residual                   8844.8   94.047
Number of obs: 302, groups: Word, 70; Speaker, 29

Fixed effects:
              Estimate Std. Error t value
(Intercept)    597.87     27.41  21.812
Mention2        -56.66     13.43  -4.220
Mention3        -55.83     13.42  -4.160
Finnf           -86.97     12.46  -6.982
log(Art.rate)  -43.86     15.03  -2.918

Correlation of Fixed Effects:
              (Intr) Mention2  Mention3  Finnf
Mention2          -0.295
Mention3          -0.201  0.505
Finnf             -0.051 -0.116 -0.133
log(Art.rt)      -0.762  0.094 -0.023 -0.155

> anova(EF_Dur.item_3)
Analysis of Variance Table
              Df Sum Sq Mean Sq F value
Mention          2 302271  151135 17.0875
Fin              1 500723  500723 56.6121
log(Art.rate)    1  75310   75310  8.5146

```

Distance to the centroid (Centr.dist)

```

(lmer(Centr.dist ~ (1|Speaker) + (1|Word), data = eng.vowels)) ->
EF_Centr.dist_0a
(lmer(Centr.dist ~ (1|Word), data = eng.vowels)) -> EF_Centr.dist_0b
(lmer(Centr.dist ~ (1|Speaker), data = eng.vowels)) ->
EF_Centr.dist_0c

> anova(EF_Centr.dist_0a, EF_Centr.dist_0b)
Data: eng.vowels
Models:
EF_Centr.dist_0b: Centr.dist ~ (1 | Word)

```

```

EF_Centr.dist_0a: Centr.dist ~ (1 | Speaker) + (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Centr.dist_0b  3 420.57 430.34 -207.28
EF_Centr.dist_0a  4 378.32 391.35 -185.16 44.244      1 2.899e-11 ***

```

```
> anova(EF_Centr.dist_0a, EF_Centr.dist_0c)
```

```
Data: eng.vowels
```

```
Models:
```

```
EF_Centr.dist_0c: Centr.dist ~ (1 | Speaker)
```

```
EF_Centr.dist_0a: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
```

```
EF_Centr.dist_0c  3 486.38 496.16 -240.19
```

```
EF_Centr.dist_0a  4 378.32 391.35 -185.16 110.06      1 < 2.2e-16 ***
```

```
Base model: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
EF_Centr.dist_0a -> EF_Centr.dist_0
```

```
(lmer(Centr.dist ~ L1 + (1 | Speaker) + (1|Word), data = eng.vowels))
```

```
-> EF_Centr.dist_1a
```

```
(lmer(Centr.dist ~ Mention + (1 | Speaker) + (1|Word), data =
```

```
eng.vowels)) -> EF_Centr.dist_1b
```

```
(lmer(Centr.dist ~ Fin + (1 | Speaker) + (1|Word), data = eng.vowels))
```

```
-> EF_Centr.dist_1c
```

```
(lmer(Centr.dist ~ log(Art.rate) + (1 | Speaker) + (1|Word), data =
```

```
eng.vowels)) -> EF_Centr.dist_1d
```

```
> anova(EF_Centr.dist_0, EF_Centr.dist_1a)
```

```
Data: eng.vowels
```

```
Models:
```

```
EF_Centr.dist_0: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
EF_Centr.dist_1a: Centr.dist ~ L1 + (1 | Speaker) + (1 | Word)
```

```
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
```

```
EF_Centr.dist_0  4 378.32 391.35 -185.16
```

```
EF_Centr.dist_1a  6 375.73 395.28 -181.87 6.592      2 0.03703 *
```

```
> anova(EF_Centr.dist_0, EF_Centr.dist_1b)
```

```
Data: eng.vowels
```

```
Models:
```

```
EF_Centr.dist_0: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
EF_Centr.dist_1b: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)
```

```
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
```

```
EF_Centr.dist_0  4 378.32 391.35 -185.16
```

```
EF_Centr.dist_1b  6 370.49 390.04 -179.25 11.832      2 0.002696 **
```

```
> anova(EF_Centr.dist_0, EF_Centr.dist_1c)
```

```
Data: eng.vowels
```

```
Models:
```

```
EF_Centr.dist_0: Centr.dist ~ (1 | Speaker) + (1 | Word)
```

```
EF_Centr.dist_1c: Centr.dist ~ Fin + (1 | Speaker) + (1 | Word)
```

```

          Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Centr.dist_0  4 378.32 391.35 -185.16
EF_Centr.dist_1c 5 378.78 395.07 -184.39 1.5395      1    0.2147

> anova(EF_Centr.dist_0, EF_Centr.dist_1d)
Data: eng.vowels
Models:
EF_Centr.dist_0: Centr.dist ~ (1 | Speaker) + (1 | Word)
EF_Centr.dist_1d: Centr.dist ~ log(Art.rate) + (1 | Speaker) + (1 |
Word)
          Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Centr.dist_0  4 378.32 391.35 -185.16
EF_Centr.dist_1d 5 375.54 391.77 -182.77 4.7853      1    0.0287 *

Model 1: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)
EF_Centr.dist_1b -> EF_Centr.dist_1

(lmer(Centr.dist ~ L1 + Mention + (1 | Speaker) + (1|Word), data =
eng.vowels)) -> EF_Centr.dist_2a
(lmer(Centr.dist ~ Fin + Mention + (1 | Speaker) + (1|Word), data =
eng.vowels)) -> EF_Centr.dist_2b
(lmer(Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker) + (1|Word),
data = eng.vowels)) -> EF_Centr.dist_2c

> anova(EF_Centr.dist_1, EF_Centr.dist_2a)
Data: eng.vowels
Models:
EF_Centr.dist_1: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)
EF_Centr.dist_2a: Centr.dist ~ L1 + Mention + (1 | Speaker) + (1 |
Word)
          Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Centr.dist_1  6 370.49 390.04 -179.25
EF_Centr.dist_2a 8 367.63 393.69 -175.82 6.8586      2    0.03241 *

> anova(EF_Centr.dist_1, EF_Centr.dist_2b)
Data: eng.vowels
Models:
EF_Centr.dist_1: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)
EF_Centr.dist_2b: Centr.dist ~ Fin + Mention + (1 | Speaker) + (1 |
Word)
          Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Centr.dist_1  6 370.49 390.04 -179.25
EF_Centr.dist_2b 7 371.75 394.55 -178.88 0.743      1    0.3887

> anova(EF_Centr.dist_1, EF_Centr.dist_2c)
Data: eng.vowels
Models:
EF_Centr.dist_1: Centr.dist ~ Mention + (1 | Speaker) + (1 | Word)

```

```

EF_Centr.dist_2c: Centr.dist ~ log(Art.rate) + Mention + (1 | Speaker)
+ (1 | Word)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Centr.dist_1    6 370.49 390.04 -179.25
EF_Centr.dist_2c   7 369.66 392.39 -177.83 2.8308      1 0.09247 .

```

```

Model 2: Centr.dist ~ L1 + Mention + (1 | Speaker) + (1 | Word)
EF_Centr.dist_2a -> EF_Centr.dist_2

```

```

(lmer(Centr.dist ~ L1 * Mention + (1 | Speaker) + (1|Word), data =
eng.vowels)) -> EF_Centr.dist_3a
(lmer(Centr.dist ~ Fin + L1 + Mention + (1 | Speaker) + (1|Word), data
= eng.vowels)) -> EF_Centr.dist_3b
(lmer(Centr.dist ~ log(Art.rate) + L1 + Mention + (1 | Speaker) +
(1|Word), data = eng.vowels)) -> EF_Centr.dist_3c

```

```

> anova(EF_Centr.dist_2, EF_Centr.dist_3a)

```

```

Data: eng.vowels

```

```

Models:

```

```

EF_Centr.dist_2: Centr.dist ~ L1 + Mention + (1 | Speaker) + (1 |
Word)

```

```

EF_Centr.dist_3a: Centr.dist ~ L1 * Mention + (1 | Speaker) + (1 |
Word)

```

```

      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Centr.dist_2    8 367.63 393.69 -175.82
EF_Centr.dist_3a  12 374.02 413.11 -175.01 1.6097      4 0.807

```

```

> anova(EF_Centr.dist_2, EF_Centr.dist_3b)

```

```

Data: eng.vowels

```

```

Models:

```

```

EF_Centr.dist_2: Centr.dist ~ L1 + Mention + (1 | Speaker) + (1 |
Word)

```

```

EF_Centr.dist_3b: Centr.dist ~ Fin + L1 + Mention + (1 | Speaker) + (1
| Word)

```

```

      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Centr.dist_2    8 367.63 393.69 -175.82
EF_Centr.dist_3b   9 368.69 398.01 -175.34 0.9455      1 0.3309

```

```

> anova(EF_Centr.dist_2, EF_Centr.dist_3c)

```

```

Data: eng.vowels

```

```

Models:

```

```

EF_Centr.dist_2: Centr.dist ~ L1 + Mention + (1 | Speaker) + (1 |
Word)

```

```

EF_Centr.dist_3c: Centr.dist ~ log(Art.rate) + L1 + Mention + (1 |
Speaker) + (1 | Word)

```

```

      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
EF_Centr.dist_2    8 367.63 393.69 -175.82
EF_Centr.dist_3c   9 366.74 395.96 -174.37 2.897      1 0.08874 .

```

Model 2: Centr.dist ~ L1 + Mention + (1 | Speaker) + (1 | Word)

```
> summary(EF_Centr.dist_2)
```

Linear mixed model fit by REML

Formula: Centr.dist ~ L1 + Mention + (1 | Speaker) + (1 | Word)

Data: eng.vowels

AIC BIC logLik deviance REMLdev

378.3 404.3 -181.1 351.6 362.3

Random effects:

Groups	Name	Variance	Std.Dev.
--------	------	----------	----------

Word	(Intercept)	0.65204	0.80749
------	-------------	---------	---------

Speaker	(Intercept)	0.21777	0.46666
---------	-------------	---------	---------

Residual		0.17079	0.41327
----------	--	---------	---------

Number of obs: 192, groups: Word, 42; Speaker, 25

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.13452	0.25302	8.436
L1ENG	0.79306	0.28663	2.767
L1NOR	0.35425	0.29633	1.195
Mention2	-0.15004	0.07306	-2.054
Mention3	-0.25707	0.07306	-3.519

Correlation of Fixed Effects:

	(Intr)	L1ENG	L1NOR	Mention2
L1ENG	-0.669			
L1NOR	-0.668	0.589		
Mention2	-0.144	0.000	0.000	
Mention3	-0.144	0.000	0.000	0.500

```
> anova(EF_Centr.dist_2)
```

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value
L1	2	1.3571	0.67856	3.9731
Mention	2	2.1345	1.06726	6.2489