Focus on English Phonetics

Focus on English Phonetics

Edited by

Biljana Čubrović and Tatjana Paunović

CAMBRIDGE SCHOLARS
PUBLISHING

Focus on English Phonetics,
Edited by Biljana Čubrović and Tatjana Paunović

This book first published 2013

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2013 by Biljana Čubrović, Tatjana Paunović and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-4925-1, ISBN (13): 978-1-4438-4925-8

TABLE OF CONTENTS

Introduction ix

PART I – Phoneme and beyond

Using MRI to see English sounds and their overlap 3
Alan Cruttenden, UK

Acquiring L2 vowels: The production of high English vowels /iː, ɪ, uː, ʊ/ by Bulgarian native speakers 13
Tsvetanka Chernogorova, Bulgaria

Original pronunciation: The accent of Shakespeare's London 27
Andrej Bjelaković, Serbia

PART II – Suprasegmentals and beyond

Pitch alignment in Welsh English – the case of rising tones in Gwynedd 45
Stefano Quaino, Austria

An acoustic analysis of the punch lines in English jokes 71
Ken-Ichi Kadooka, Japan

Observations on the nucleus in English and Serbian 85
Brian Mott, Spain

Methodological issues in the acoustic analysis of spontaneous speech prosody 99
Aleksandar Pejčić, Serbia

The status of intonation in a level approach in the organization of language 115
Vladimir Phillipov, Bulgaria

Intonation patterns and phonetic stereotypes: New life for old terminology 133
Yulia Nenasheva, Russia

Intonation interference and its impact on effective communication between native/non-native speakers 151
Oksana Pervezentseva, Russia

PART III – Applied phonetics and beyond

To flip or not to flip? Phonetics, phonology and the flipped classroom 165
Patricia Ashby, UK

Minimal pairs in English phonetics teaching 183
Rastislav Šuštaršič, Slovenia

Beginnings, endings, and the in-betweens: Prosodic signals of discourse topic in English and Serbian 191
Tatjana Paunović, Serbia

British or American pronunciation? 215
Snezhina Dimitrova, Bulgaria

Slavic English accents revisited - A case study of Russian Serbian- English in films 227
Biljana Čubrović, Serbia

Phonetic similarity in transliterated English trademarks – A preliminary study in Japanese 237
Isao Ueda, Japan

PART IV – Phonology and beyond

Level ordering of affixes: a phonological perspective 251
Jelena Vujić, Serbia

The functional classification of English vowels: Phonological and orthographic evidence 261
Csaba Csides, Hungary

Contributors 283

Index 291

INTRODUCTION

Focus on English Phonetics is the third collection of papers created by scholars gathered around the *Belgrade International Meeting of English Phoneticians*, started in 2008 by Professor Biljana Čubrović of the Faculty of Philology, University of Belgrade. After *Ta(l)king English Phonetics Across Frontiers* (2009, CSP) and *Exploring English Phonetics* (2012, CSP), this collection represents a further step in the same direction. As the ones before it, this volume aims to bring together researchers in the rich field of English phonetics, and provide them with a forum for exchanging ideas and research experience. The 18 contributors to this volume come from different linguistic and academic backgrounds, and from 9 different countries: Austria, Bulgaria, Hungary, Japan, Russia, Serbia, Slovenia, Spain and the United Kingdom. As a result, the volume reflects the authors' diversity by both its breadth and tenor. The topics discussed, the research approaches used, and the variety of theoretical, applied and experimental aspects of phonetic investigations all speak of this diversity, a very desirable quality in any field of research.

What this collection also shares with the two previous volumes is its general outlook and organization. The chapters are organized into four thematic units. *Part One* deals with segmental issues, *Part Two* with questions of suprasegmental phonetics. *Part Three* looks into some issues related to English phonetics and phonology teaching and EFL pronunciation instruction, while *Part Four* turns to some questions related to English phonology.

Alan Cruttenden's chapter, titled *Using MRI to see English sounds and their overlap,* opens the first part of the book by a description of the use of Magnetic Resonance Imaging in the investigation of segmental articulation and, particularly, coarticulation phenomena. The author discusses the problems in the instrumental study of articulation, particularly the problem of observing the dynamics of articulatory movements in speech production, where single static images produced by x-ray based technology were not very helpful. Cruttenden describes the introduction of Magnetic Resonance Imaging into articulation study, particularly a new procedure recently developed at Oxford, which can produce dynamic images showing the tongue movements during articulation. Such images have been examined in a number of phrases containing examples of

different English phonemes, available to the public on the website attached to the 7th edition of *Gimson's Pronunciation of English.* The author illustrates his description by a number of MRI images used in this procedure.

The acquisition of English monophthongs by Bulgarian EFL learners was investigated by Tsvetanka Chernogorova. Her chapter, titled *Acquiring L2 vowels: The production of high English vowels by Bulgarian native speakers*, presents the results of a research study which focuses on Bulgarian learners' problems in acquiring the complex vowel system of English. Compared to the English system of twelve monophthongs, the Bulgarian vowel system is rather simple, containing only six vowels, which makes EFL students' acquisition very difficult. The study particularly focuses on the English high vowels /iː, ɪ, uː, ʊ/, which the author describes as especially difficult since they are all, at least seemingly, similar to Bulgarian high vowels /i/ and /u/. The study explores the vowel production of first-year English majors at the University of Sofia, who read lists of words in citation form in English and Bulgarian. The analysis presents acoustic formant measurements (F1, F2) and the results of statistical analyses. Different degrees of acquisition of high English vowels by Bulgarian students are noticed, from complete substitution by the L1 vowel, e.g. /i/, to various degrees of quality alteration towards L2 vowel quality or, in some cases, very high degrees of acquisition.

The diachronic perspective in phonetic investigations is illustrated by Andrej Bjelaković's chapter, titled *Original pronunciation: The accent of Shakespeare's London.* The author provides a sketch of Early Modern English (EME) pronunciation, based on several major works dealing with EME phonology, illustrating the main differences in pronunciation between the present-day English and that of the 17th century. The chapter is organized around individual vowel phonemes, with the standard lexical sets of contemporary English used as the starting point, and with illustrations, in the form of phonemically transcribed verses, taken from William Shakespeare's works. The author particularly focuses on the frequently quoted observation that the phonological system of EME was rather similar to present-day English system, while the differences lie in the phonetic realisation. He also illustrates the point that mainstream EME pronunciation share many features with some regional accents of contemporary English.

Part Two of the book, devoted to suprasegmental topics, opens with a discussion of intonation in Welsh English. Stefano Quaino's chapter, titled *Pitch alignment in Welsh English – the case of rising tones in Gwynedd*,

sets off from the common observation that Welsh English and Celtic English are well-known for their preference for rising tones in declaratives, analysed by several authors (Gimson 2008; Cruttenden 1995; Tench 1990; Walters 1999). Focusing particularly on the previous observations that rises may be the most frequent tone on declaratives (Gimson 2008:289), and that the increase in the use of rises on declaratives may be 'a strong Celtic influence' (Cruttenden 1995:155), the author presents the results of his own acoustic analysis of Gwynedd English, aiming to describe and explain its main prosodic characteristics with respect to the use of rising tones.

An acoustic analysis of the punch lines in English jokes is presented in the chapter by Ken-Ichi Kadooka. The author uses acoustic analysis to explore the description of the punch-line paratone in jokes as a subtype of the low paratone, characterized by a combination of phonetic features, such as a lower pitch, slower tempo, and a pause preceding it, by which the end of the joke is signalled. The punch line paratone also includes a gradual lowering of the baseline pitch from the beginning towards the end of the joke - until the punch line. The author presents an analysis which illustrates this description, to a certain extent.

Brian Mott presents a comparative study of English and Serbian intonation, specifically the characteristics of the nucleus. In the chapter titled *Observations on the nucleus in English and Serbian* the author summarizes some differences between English and Serbian utterances as regards the position of the nuclear stress. The versatility of the nucleus in Serbian was established through a number of informants, and then a list of examples was composed, aiming to test which elements of the Serbian sentence can receive tonic stress, and what concomitant changes in their pragmatic value this would entail. The author analyses the recordings and classifies the utterances into a number of types, such as WH-questions, interrogatives with an emphatic particle in Serbian, and those containing negative adverbs, intensifiers, emphatic pronouns or possessives, in order to compare English and Serbian in this respect.

Methodological issues in the acoustic analysis of spontaneous speech prosody are discussed by Aleksandar Pejčić. The author presents the methodological design of a research study investigating the prosodic characteristics of Serbian and British persuasive political speech, and uses this example to discuss some methodological problems and difficulties common to most spontaneous speech prosody studies. The author especially highlights the problems related to the choice of suitable speech tokens, in terms of their subject, register, style, as well as the regional, gender and age differences of the speakers. The problems include the

variety of sources used for the extraction of speech tokens, as well as the need to address speech errors and repairs made by speakers.

Vladimir Phillipov discusses a model of intonation representation in his chapter *The status of intonation in a level approach in the organization of language*. The discussion sets off from the traditional representation of intonation between phonology and syntax, and the recent generativist suggestion that intonation conveys postlexical pragmatic meanings, or, in the author's words, that "it occupies the 'safest' component that comes hierarchically after syntax". The author presents a view of intonation as an exponent of fluctuation, or a shift in the status of a linguistic item leading to a different function, while preserving the form. The analysis attempts at establishing a correlation between syntax and intonation.

In the chapter by Yulia Nenasheva, titled *Intonation patterns and phonetic stereotypes: New life for old terminology* different approaches to prosodic research are discussed, and some results of intonation study presented. To illustrate the point that the meaning of an utterance is expressed through the arrangement and interaction of prosodic elements in an intonation pattern, Nenasheva presents a study of the prosodic components of the utterance, such as its durational, dynamic and tonal qualities. The analysis involves acoustic measurements and a statistical analysis, which indicate that prosodic elements comprise a complex structure of interrelated units, that the arrangement of the units is predictable, and that it carries specific meanings. The research study shows that these prosodic complexes possess certain distinctive features, and that sets of these features identify them as intonation patterns that serve as models in speech production, through a realization of phonetic stereotypes.

Oksana Pervezentseva looks into the topic of *Intonation interference and its impact on effective communication between native/non-native speakers.* The study the author presents focuses on the ways in which prosody affects communication between native and non-native speakers in situations of artificial bilingualism, specifically the communicative-pragmatic types of utterances that are likely to be subject to L1 interference, and to cause miscommunication. The research presented is based on the zone conception of intonation, and the findings indicate native speakers' sensitiveness to the inaccurate use of intonation patterns, mostly in the emotional-modal aspect.

Although some of the papers in the first two parts of the volume also include English phonetics from the perspective of non-native speakers, *Part Three* is specifically devoted to issues of 'applied phonetics', either

phonetics and phonology teaching or pronunciation training as part of EFL learning and teaching.

Patricia Ashby opens this part by a description of a new teaching methodology, labelled 'the flipped classroom'. In her chapter titled *To flip or not to flip? Phonetics, phonology and the flipped classroom* Ashby describes the first steps in introducing modern technologies such as screen-capture software, or educational vodcasting (videocasting), which aim at enhancing students' learning experience, increase their motivation, and create an environment in which they would develop their potentials more fully. The author describes how the success of these first steps encouraged Bergmann and Sams to introduce the technique of the flipped classroom (Sams 2010; Bergmann & Sams 2012), as the 'asynchronous online delivery of lectures', or switching the place of homework and lectures, so that vodcast lectures are watched at home before the class, while class time is freed for hands-on work, various activities or discussion. The author describes the use of the 'flipped classroom' in phonetics or phonology teaching and presents the results of a study of the effectiveness of this technique in a course attended by final year students of phonology at the University of Westminster. The study compares success rates (grades) in a traditional group or 'cohort' to those of a 'flipped cohort' achieved over two weeks of a 12-week course. The author concludes that 'flipped lectures' indeed result in students' greater confidence and deeper knowledge of the subject-matter.

Minimal pairs in English phonetics teaching are discussed by Rastislav Šuštaršič, from the point of view of English-Slovene contrastive analysis and teaching of English pronunciation. The author states that a first step in this process should be the identification of the phonemic contrasts in English, their frequency of occurrence, and then their possible application in specific pronunciation classes. The author describes the main differences between the sound systems of English and Slovene, so that the practice can focus on those distinctive sounds which are particularly problematic for Slovene students of English. He illustrates his point by an inventory of minimal pairs taken from John Higgins, and some possible approaches and activities based on involving minimal pairs in pronunciation classes.

Tatjana Paunović discusses the use of prosodic cues at the discourse level, specifically, in signalling discourse structure. In the chapter titled *Beginnings, endings, and the in-betweens: Prosodic signals of discourse topic in English and Serbian,* the author presents a research study which investigates how prosodic cues are used to signal discourse topic beginning, continuation and ending in a reading task performed by two

groups of participants: L1 speakers of Serbian, who are also EFL learners, and L1 speakers of British English. The acoustic analysis includes F0/pitch, intensity, and duration measured at intonation unit boundaries, first peak/onset, and nuclear accent syllable, and overall pitch range and intensity of intonation units. The statistical analysis points to some important differences between the native-speaker group and the EFL group in reading the English text, while certain similarities, as well as differences were identified in the English and Serbian texts when read by their respective native speakers. The author points out that some, but not all of the EFL students' problems in reading the L2 text could be attributed to L1 prosodic transfer.

Snezhina Dimitrova's chapter, titled *British or American pronunciation?* turns to the issue of language attitudes among Bulgarian EFL learners. The study she presents compares the students' pronunciation preferences with their spoken performance, based on analyses of the students' recordings. The analysis of forty-seven recordings involved auditory and acoustic analyses, aiming to establish how consistent the Bulgarian tertiary-level learners were in their use of the well-known salient segmental and suprasegmental features of the pronunciation model of their choice. The study shows that the vowel quality of words from the LOT and BATH lexical sets, along with rhoticity and t-voicing, position of lexical stress and variable individual word pronunciations are among the most prominent traits that students use inconsistently when trying to imitate the British Received Pronunciation or the General American accent.

Biljana Čubrović's study deals with the topic of linguistic credibility of what film industry approves of as acceptable Russian accents. The chapter entitled *Slavic English accents revisited - A case study of Russian Serbian-English in films* provides the results of the phonetic analysis of Rade Šerbedžija's speech in four recent films. Some segmental phonetic features of the actor's English idiolect are studied with the aim of establishing how much effort was invested into him sounding like a Russian native speaker, and whether he can pass as one. Carefully selected audio recordings are analyzed from a segmental viewpoint, with the help of acoustic phonetics tools, and also from an auditory perspective, where necessary. The results of the analysis show that the actor's "Russian-coloured" speech does not necessarily include the most striking features of Russian EFL speech like palatalization, but his South Slavonic language background seems to be satisfactory for an international audience.

The chapter titled *Phonetic similarity in transliterated English trademarks – A preliminary study in Japanese* by Isao Ueda investigates

the phonetic and phonological problems encountered by applicants for new foreign trademarks, related to relevant legal decisions. Ueda explains the trademark and application procedure in Japan, and describes a number of examples of trademarks which were turned down by Japanese examiners because they were judged to be phonetically similar, although they sounded (and looked) completely different to the foreign applicants. The author states that the main reason for this apparent discrepancy is the Japanese trademark law, which demands newly proposed foreign trademarks to be transliterated into Japanese orthography. This process can result in a kind of phonetic distortion of the original form. The author discusses several examples of trademarks in which the similarity decision was affected by certain factors, such as segmental contents, the different position of the syllable in a word, and the total length of the trademark. Some possible improvements of the procedure are suggested, based on an example of apparent inconsistent judgment.

Part Four of the collection turns to two phonological issues. The contribution by Jelena Vujić, titled *Level ordering of affixes: A phonological perspective*, examines the phonological aspect of the mechanisms responsible for affix combinations in English, particularly when it comes to level ordering and restrictions that govern affix combinations in English. To illustrate the point that phonology and morphology interact closely, the author offers an outline of various theoretical approaches to English word-formation, from TG, via Siegel's views (1974), Kiparsky's Lexical Phonology and Morphology model (Kiparsky 1982), or Giegerich's views (1995), to the most recent Optimality Theory (Raffelsiefen 2004). The author points out that all these different theoretical approaches acknowledge some aspects of the interdependency of phonology and word-formation.

Csaba Csides closes the volume by the chapter titled *The functional classification of English vowels: Phonological and orthographic evidence,* which focuses on the phonological evidence and orthographic justification underlying the division of English vowels into tense and lax. Csides points out that in addition to phonetically-based explanations, namely, that tense vowels are produced with more tension of the articulatory muscles, more length and a degree if diphthongization, phonological alternations seem to support the view that the categories of tense and lax are indeed functional/phonological in nature. Phonological processes discussed in connection with these arguments are Vowel Shift, Trisyllabic Laxness, Laxing by ending, CiV laxing, Pre-cluster laxing and Laxing by free U. In the second part of the chapter, Csides focuses on the regular sound values of English vowel letters and discusses the difference between free and

covered graphic positions. Based on orthographic evidence, the author concludes that tense and lax vowels respectively tend to occur in different types of graphic (orthographic) positions in the default case, but also that the effect of the free position rule may be eliminated by overriding regularities that are phonological in nature.

* * *

Focus on English Phonetics presents empirical research findings, and can, therefore, be of special interest to other researchers in the old but nonetheless exciting field of phonetics. It can also be appealing for graduate and doctoral linguistics students, since chapter authors also discuss some theoretical questions, models of representation and recent methodological approaches. Lastly, we believe that the contributions in this volume, although not all of them deal with teaching and learning problems, can also help applied phonetic practitioners or EFL teachers, since some of the topics discussed stem from extensive classroom experience and the problems observed working with EFL or phonetics students. Therefore, we hope that this collection will find a way to communicate to an audience at least as diverse as the authors and topics of these chapters.

Editors
April, 2013

PART I.

PHONEME AND BEYOND

USING MRI TO SEE ENGLISH SOUNDS AND THEIR OVERLAP

ALAN CRUTTENDEN

Outline

Until the 20th century we had no way of looking directly at tongue movements. In the 20th century x-rays were used to study such movements. At first making x-rays was too expensive for this procedure to be used very often. As the procedure became cheaper, a new problem arose: x-rays were thought to present a cumulative radiation hazard to health if used too frequently. So only a very limited number of x-rays of the tongue were ever made and of course those that were made showed only single static images, so the subject would have to hold an articulation for an unnatural length of time. With the introduction of the new procedure of Magnetic Resonance Imaging the radiation hazard has been overcome. But this could still produce only single static images. A new procedure has recently been developed at Oxford whereby we can produce dynamic images actually showing the tongue moving in the mouth. Such images are examined in a number of phrases containing examples of all the phonemes of English. The dynamic images are available on the website attached to the 7th edition of *Gimson's Pronunciation of English*.

1. Introduction

A new method of studying the articulation of vowels and consonants has been developed at the Phonetics Laboratory in Oxford. Most people have heard of Magnetic Resonance Imaging or MRI: it is one of the successors to the old-fashioned X-ray. It uses changes in electro-magnetic current to study various tissues of the body. It has been extensively used over the last twenty years for looking at various internal organs of the body, e.g. the heart and the liver. I, for example, had MRI images made of my lumbar spine twenty years ago.

More recently there has been a development called dynamic MRI: this means producing pictures of movements as opposed to states. Dynamic MRI has been used to picture blood flow through the brain and heart. Producing single MRI images is very slow, much too slow on its own to show the rapid movements involved in, say, blood flow through the heart. Therefore, use is made of the fact that blood flows through the heart in regular pulses (i.e. the heart beat). To capture regular movement, an MRI image is taken at a slightly later stage in each pulse; when these images are put together, they show the continuous movement involved in each pulse.

A method has now been developed using a similar technique to study the movement of the tongue in the mouth. To make MRIs of tongue movements an element of repetition has to be built in (like the pulses of blood in the heart). Individual speakers are trained to repeat a short phrase (like *ride in fog*) rhythmically to the beat of a metronome. MRI images are produced slightly later on each repetition. Putting the images together produces a picture of the tongue movement throughout the phrase.

One advantage of these MRIs over X-rays (as well as the absence of the radiation hazard) is that they show soft tissue variations more clearly. The disadvantage is that some hard tissue does not always show up very well; for images of the mouth this refers mainly to the upper teeth. Additionally these MRIs are taken in only one plane, the midsagittal section, with the result that vertical movement is shown but horizontal movement is not, so the larynx can be seen moving up and down while the opening and closing of the vocal cords cannot be seen. There is thus no direct information on voicing.

2. Gimson's Pronunciation of English

For the website accompanying the seventh edition of *Gimson's Pronunciation of English*, dynamic MRIs of fifteen English phrases were produced by a speaker chosen as representing a modern form of RP (a female in her twenties). The phrases were chosen to include all the consonants and vowels of English. MRI is an expensive procedure and my fifteen phrases were a free addition to the end of an experiment which had its own funding. If I were able to collect the MRIs again, there are some things I would change. I would, for example, change the angle at which the images of the head and mouth were taken; you will see that the head is tilted backwards because of the way the informant was lying. I would also make some of the phrases a little more realistic (a phrase like *Crawl a zoo* is rather peculiar!). But neither of these things affects the basic validity of

the videos. To view the videos the website has to be accessed at http://www.hodderplus.co.uk/linguistics and *Gimson's Pronunciation of English* selected from the sidebar. At this point registration is necessary, although it is free and purely for contact purposes. After registration, the list of MRI videos (in wmv) can be seen; select the one entitled *Ride in fog*. It can be downloaded as a web archive if so wished. Select the arrows on the bottom bar to go to full screen view and use the slider or the space bar to start and stop.

3. Ride in fog

The various parts visible can fairly easily be identified. Firstly the vertebrae and the intervertebral disks can be seen, with the spine itself behind them. In front of the vertebrae is the trachea (the black tube) and, on a level with the bottom of the chin, the larynx. Just in front of the point where the spine goes into the head can be seen the uvula; between the uvula and the spine a gap will open up in the nasopharynx during the production of nasals. In front of the uvula and along the top of the mouth can be seen the soft palate (relatively bright-looking) and the hard palate and the teeth ridge (not so bright-looking) and then the lips. Notice again that we cannot see the upper teeth. Filling almost the whole of the mouth is the ball of the tongue.

Now, look at *Ride in fog* in more detail. You can use the bottom slider to advance the video slowly. In the first eighth you can see the tip of the tongue somewhat curled to the rear of the teeth ridge for [ɹ]. Around one quarter the body of the tongue is in position for the first half of the diphthong [ä] and, approaching one third, the front of the tongue moves towards the hard palate for the second half of the diphthong [ɪ]. The tip of the tongue is now forming the closure for the [d]. You can see the tip of the tongue showing up as the bright bit; although not able to see the teeth, you can judge that the tip is against the juncture between the teeth ridge and the upper teeth. There is no realization of the [ɪ] between the [d] and the [n]; rather the nasopharynx can be seen opening behind the uvula to change oral [d] into nasal [n]. This can be seen around half way. The next movement (as the nasopharynx closes again) shows the lower lip moving upwards to make contact with upper teeth for the [f] (the lip movement but not the upper teeth showing for the labiodental articulation). By two thirds the base of the tongue is pulled backwards to form the lower back vowel [ɒ] and following this the back of the tongue then rises up to make a closure for the [g]. Notice that, even following a low back vowel, the

closure is still in the middle of the soft palate and nowhere near being uvular.

4. Curious beer

Now select the MRI video *Curious beer*. The tongue starts at the position for /k/, similar to the position for /g/ in the previous *fog* but further forward because it is before a [j]. Around one eighth of the phrase the body of the tongue has slid forward for the front position for [j]. The [ʊ] is hardly visible on the video; it is obviously even more fronted than usual because of the preceding [j] and it might be seen (from a front view) in some rounding of the lips which of course is not visible in the video. The following [ɹ] shows no curling of the tip (as it had at the beginning of *ride*) but is made by raising the blade of the tongue towards the back of the teeth ridge; evidently this is like the 'bunched [ɹ]' often described as a common realization of /r/ on American English: see, for example, Zhou *et al.* (2008). The [ɪə] moves the front of the tongue back while remaining raised and then lowers it somewhat. At the same time the tip is moving forward and at around half way is at the base of the teeth ridge for [s]. The lips now move together for the [b], the front rises again for the [ɪ] and finally lowers somewhat again for the [ə].

5. Pain in the mouth

As a final illustration I will do a walkthrough of the video *Pain in the mouth*. The opening picture shows the lips closed. The lips open and the front of the tongue rises through the [e] position to the [ɪ] position. Then the tip can be seen rising to touch the teeth ridge (around a quarter way through). Notice that at the same time the nasopharynx has opened. There is no [ɪ] between the <n> in *pain* and the <n> in *in*, just a long [nː]. Just before half way through the video the tongue tip is moving from its position on the teeth ridge to a further forward position on the teeth for [ð]; notice that the nasopharynx is still open through this dental fricative. The body of the tongue then moves back to a neutral position for [ə] while the lips come together for the [m]. At the end of the [m] the nasopharynx closes. As the lips open the body of the tongue lowers for the [ä] before the back rises for the [ʊ] while the tip is moving towards the teeth again for the [θ].

6. Articulations of /r/

I now consider what the videos tell us about one particular articulation, that for /r/. Recall the position for /r/ in *Ride in fog,* shown in Figure 1. The /r/ here is before a front (or mid) open vowel. It has the tip of the tongue curled back to a position at the back end of the teeth ridge; in fact it looks from the video as if it is almost touching it.

In *Crawl a zoo* the articulation of /r/ is much the same, although overall the articulation shows the position before a back vowel, as in Figure 2.

But in *Curious beer* the /r/ involved a raising of the front and blade of the tongue, as shown in Figure 3. In this phrase /r/ is made like this presumably because it is between close /ʊə/ (which is monophthongized and almost fronted to [ÿː] following /j/) and close beginning of /ɪə/.

Similarly in *Dream of debt,* as shown in Figure 4. Here /r/ follows the alveolar /d/ and precedes the front vowel /iː/. Hence curling the tongue tip back would be a rapid contortion which is avoided by instead bunching the very front of the tongue up towards the front of the hard palate.

Figure 1. /r/ in *Ride in fog*



Figure 2. /r/ in *Crawl a zoo*



Figure 3. /r/ in *Curious beer*



Figure 4. /r/ in *Dream of debt*



Similar variation in the realizations of /r/ in American English have been previously shown (Delattre & Freeman 1968; Westbury, Hashi & Lindstrom 1998; Zhou *et al.* 2008; Stavnes *et al.* 2012) but it is apparent that such variation also occurs in this speaker of British English. Zhou et al. refer to this articulation as 'bunched' although it is not clear how 'bunching' applies as opposed to just raising.

7. Velars, close back vowels, and laterals

Some other interesting articulations (confirming or provoking current thinking) are shown on the MRI videos on the *Gimson* website:

(a) The velars /k, ɡ, ŋ/ are seen as generally farther forward than is usually imagined. Even before the open back vowel /ɑː/, as is seen in *Guard my thumb*, /ɡ/ is articulated in the middle of the soft palate and certainly nowhere near approaching the uvula. Following [ɪ] in *A weird thing* /ŋ/ is towards the front of the soft palate (though not on the hard palate). As expected, before /j/ in *Curious beer* /k/ is on the verge of being palatal.

(b) Many recent descriptions of RP, e.g. Hawkins and Midgley (2005), Cruttenden (2008: 81), note the recent fronting of the /uː, ʊ, ʊə/. This can be clearly seen in the videos containing these vowels. The only example seen in the phrases I have discussed here is in *Curious beer* where of course co-articulation following palatal /j/ predictably produces fronting of /ʊə/. But in other phrases on the website it is also very apparent, e.g. /uː/ in *zoo,* in *July* (no matter whether the vowel is analysed as /uː/or /ʊ/), and /ʊə/ in *tour*.

(c) No /l/'s figured in the current discussion but the videos do show noticeably different tongue positions for clear and dark /l/: varieties of clear [l] can be seen in *valley, leisure, line, July,* and *curly*; and varieties of dark [ɬ] in *crawl* and *bull.* Neither of the last two words showed vocalisation of dark [ɬ], possibly because the speaker was conscious of articulating relatively carefully.

8. Conclusion

The procedure for producing 'dynamic' videos using repetition is artificial (it is described as only an 'animation' on the website of the Phonetics Laboratory of the University of Oxford—see references below). In various laboratories in the U.S., experiments are being made with MRI scanners which reduce the interval between scans (Zhou *et al.* 2008) so that real-time movements can be scanned but both the instrument and expense still put limits on what can be done. At the moment, real-time video MRI can only operate between 3 and 6 frames a second whereas the constructed videos on the Gimson website have around 34 frames a second. No doubt in time this will be overcome.

MRI video has the potential for confirming much articulatory information that we previously knew (if at all) only by our feeling for what is going on in the mouth. This particularly applies to allophonic variation and gestural overlap. We have other equipment (like palatography) which can measure static positions but MRI leads in portraying movement.

The principal worth of these videos to teachers is to better inform their own knowledge. But it can also bring home to students the continual change and overlap in articulation and get them away from the idea of static sequences which can be an unfortunate side-effect of transcription. It can of course also be another way to spark an interest in phonetics.

References

Cruttenden, A. 2008. *Gimson's Pronunciation of English.* Seventh edition. London: Hodder Education. Related videos retrieved 17th April 2012 from http://www.hodderplus.co.uk/linguistics.

Delattre, P. & D. C. Freeman. 1968. A dialect study of American English r's by x-ray motion picture. *Linguistics* 44: 28–69.

Hawkins, S. & J. Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35: 183–199.

Phonetics Laboratory, University of Oxford. Magnetic Resonance imaging of the moving vocal tract. Retrieved 17th April 2012 from http://www.phon.ox.ac.uk/mri.

Stavness, I., B. Gick, D. Derrick, and S. Fels. 2012. Biomechanical modality of English /r/ variants. *Journal of the Acoustical Society of America* 131: 355–360.

Westbury, J. R., M. Hashi & M. J. Lindstrom. 1998. Differences among speakers in lingual articulation for American English /ɹ/. *Speech Communication* 26: 203–226.

Zhou, X., C.Y. Espy-Wilson, S. Boyce, M. Tiede, C. Holland & A. Choe. 2008. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *Journal of the Acoustic Society of America* 123: 4466–4481.

ACQUIRING L2 VOWELS: THE PRODUCTION OF HIGH ENGLISH VOWELS /iː, ɪ, uː, ʊ/ BY BULGARIAN NATIVE SPEAKERS

TSVETANKA CHERNOGOROVA

Outline

Unlike the English language, which has 12 monophthongs, Bulgarian has only six. This difference in number makes the acquisition of L2 vowels difficult for Bulgarian learners of English, who find the production of the English high vowels /iː, ɪ, uː, ʊ/ especially difficult, because of their seeming similarity to the Bulgarian high vowels /i/ and /u/. This paper presents the results of a study of the production of L2 high vowels and attempts to show how the articulation of L1 high vowels interferes in this process and affects the acquisition of L2 high vowels. The subjects of the study were first year students at the University of Sofia whose major is English philology. They were recorded pronouncing a number of English words in their citation form containing the high vowels /iː, ɪ, uː, ʊ/ and a number of Bulgarian words containing the high vowels /i/ and /u/. Spectral qualities (F1 and F2 values) were measured and statistically analysed. The results show different degrees of acquisition of L2 high vowels – from complete substitution with the L1 high vowel /i/, through an acceptable alteration of the quality of the sound in the direction of L2 to satisfactory or very good acquisition.

1. Introduction

Different aspects of the contrastive phonetics and phonology of English and Bulgarian have already been examined at various length in a number of previous studies (Minkoff 1973; Despotova 1978; Danchev 1988, 1990) which have raised numerous issues regarding L1 transfer in the production and perception of L2 sounds. Due to fundamental

differences between the vowel systems of Bulgarian and English, Bulgarians tend to disregard the qualitative-quantitative distinction between the English long and short vowel phonemes. Bulgarian learners of English find the production of the English high vowels /iː, ɪ, uː, ʊ/ especially difficult, because of their seeming similarity to the Bulgarian high vowels /i/ and /u/. Thus, it can be said that the process of acquiring L2 phonological categories is greatly impeded by the categories of one's L1. The results of some studies have shown that, in establishing the new categories in the foreign language, non-native speakers may resort to phonetic cues from the native speakers (Bohn & Flege 1992; Flege *et al.* 1997; Escudero 2002).

The purpose of this study is to investigate how the acquisition of phonological contrasts and phonetic realizations in a second language is affected by the first language phonological knowledge. Although initially the study did not aim to test the two models of L2 vowel acquisition, Flege's *Speech Learning Model* – SLM (Flege 1995) and Best's *Perceptual Assimilation Model* – PAM (Best 1995), the results of the experiment prompted the author to make these two the most influential models which explain the perception and production of L2 phoneme categories in relation to the categories found in the learners' native language part of this paper.

Flege's model predicts that L2 vowel phonemes which exhibit sufficient phonetic difference from the L1 target categories are developed earlier into new phonetic categories by non-native speakers. Best's model describes a process by which we perceptually assimilate non-native phonemes into our own phonemic inventory. If a foreign category is similar to an L1 category, it will be assimilated to it. Therefore, if a single L1 category is similar to two distinct L2 sounds, discrimination is expected to be poor. So these two models are also tested and it is shown that the degree to which an L2 sound is acquired greatly depends on its articulatory and auditory similarity to or difference from the corresponding L1 category.

This paper presents the results of an experiment which investigates the production of the English front and back high vowels /iː, ɪ, uː, ʊ/ by a group of first year students who study English philology at the University of Sofia in an attempt to show how the production of L2 vowels by relatively proficient speakers whose first language is Bulgarian is influenced by the prior L1 phonological knowledge and to show that the extent to which L2 sounds have been acquired depends on the articulatory and auditory similarities and differences between the two languages.

The paper begins by giving a brief overview of the English and Bulgarian vowel systems making an attempt to show the potential difficulties that Bulgarian speakers encounter while mastering the English high vowels. The methodology is presented in section 3 and the results are shown and analysed in section 4, followed by discussion. The final section of the paper includes general conclusions and some methodological suggestions for overcoming the observed difficulties in the production of L2 vowel categories.

2. English and Bulgarian vowels

The vowel system of English is relatively large and complex. Most languages have between five and seven vowels, but English has at least twenty, twelve of which are monophthongs or "relatively" pure vowels (Cruttenden 2008: 98). The phonetic description of the vowel sounds in English, i.e. the way they are pronounced, is used to classify them into three important groups of vowels: the short vowels, the long vowels and the weak vowels. English vowels are also traditionally divided into tense and lax categories. The vowels produced with the front of the tongue raised as high and as close to the palate as possible such as /iː, ɪ/, are high (close) front vowels and the back high vowels /uː, ʊ/ are produced with the back of the tongue in the same position. Other vowels produced with the back of the tongue are /ɑː ɔː ɒ/ and depending on the degree of raising of the tongue they are low (open)-/ɑː/ because of the considerable distance between the tongue and the roof of the mouth, high-to-mid /ɔː/ and low-to-mid /ɒ/ when the tongue is in between these two extremes. The other vowels occupying the front region of the traditional vowel quadrilateral are mid /e/ and mid-low /æ/. For the central vowels /ʌ ɜː ə/ the highest point of the tongue lies in an intermediate position at the junction of the hard and the soft palates and depending on the distance to the roof of the mouth they are mid vowels /ɜː, ə/ and mid-low /ɑː/.

The vocalic system of standard Bulgarian is relatively simple. It has no distinction based on phonological quantity, i.e. there are no long and short vowels. All vowels are relatively lax as in most other Slavic languages. The six Bulgarian vocalic phonemes are evenly distributed in the vowel space and are classified according to three distinctive features:

- front, middle and back
- rounded and unrounded
- high, mid, and low.

The front high vowel is /i/ and /e/ is front, open-mid and they are unrounded. The backmost vowels, /u, ɔ/, are rounded, /u/ is high and /ɔ/ is open-mid. /ɤ/ is near-back and is unrounded. The last vowel phoneme is /a/, the tongue is in a normal (middle) position, neither lifted up nor moved forwards or backwards. So this vowel is classified as middle, low and unrounded. /a/ and /ɤ/ are neutralised in unstressed positions and we get another vowel sound [ɐ]. Similarly, /o/ is the unstressed neutralisation of /ɔ/ and /u/.

The above brief description of the vowel systems of the two languages shows the existence of a number of important differences which inevitably create conditions for L1 transfer in the production of the English vowels by Bulgarian native speakers of English.

3. Methods and subjects

The subjects of the study are 10 female first year students of English philology at the University of Sofia. All subjects do a course in practical phonetics and are approximately the same age (18–20).

The subjects were recorded pronouncing a number of English words in their citation form containing the high vowels /iː, ɪ, uː, ʊ/ and a number of Bulgarian words containing the high vowels /i/ and /u/. [1] When creating the list of words care was taken to choose words in similar phonetic contexts in the two languages. All the words with three exceptions are of the pattern CVC, and for the English words, the consonant in final position is always voiced and different from /l/ to avoid any possible influence on the quality and quantity of the vowels studied by the sound /l/ or by pre-fortis clipping. Also, words in which the vowels occurred after the approximants /j/, /w/ and /r/ were avoided, as the approximants would have severe coarticulatory effects on the locations of the first two formants. However, for the back high vowels /ʊ/ and /uː/, it was very difficult to avoid these environments, so this reflected on the number of words. There are eight words for each of the front high vowels in English

[1] The list comprised the following words: *eager, bee, lead, leave, mean, flee, heed, beam, bid, lid, big, tin, fin, sin, hymn, chin, book, took, foot, put, push, move, moon, tomb, zoo, pooh*. The Bulgarian words were: *nit, kit, mit, chin, sin, vid, bit, dim, sit, hit, um, but, suma, chuk, runo, lud, kum, rus*.

/ɪ, iː/ and for the Bulgarian /i/ vowel, and five words for the high back vowels respectively.

The recordings were made in a soundproof language lab of the Faculty for New and Classical Philologies at the University of Sofia. They were digitalised at a sampling rate of 44.1 kHz and analysed using Praat 4.3.36 program for speech analysis (Boersma & Weenink 2005). The formant frequencies of F1 and F2 were measured at a point of the vowel where formant frequencies were judged to be relatively steady and representative. The mean values of each formant were calculated and the data on standard deviation was obtained for each speaker. The control values of formant frequencies of vowels produced by native speakers of English, used as the basis for comparison in the study were taken from the data on formant frequencies in RP published by Deterding (1990, 1997). No control values of formant frequencies for the Bulgarian vowels are used; the study uses the formant frequencies of the subjects' production of the L1 vowel sounds and compares them to the formant frequencies of the subjects' production of the L2 categories.

The mean values obtained were used as the basis of plotted graphs of F1 and F2 in which the comparison of native and non-native production was visually represented.

4. Results

4.1. Front vowels

Bulgarian has only one vowel in the close front region. In terms of quality, it is closer to /iː/ than to /ɪ/. In terms of length, it is shorter than /iː/ but usually longer than /ɪ/. Therefore, both the quality of the English short /ɪ/ and the length distinction between /iː/ and /ɪ/ pose problems for Bulgarian learners of English. Several predictions can be made regarding the potential difficulties and transfer from L1 in the subjects' production of the two front English vowels. Firstly, since it is very similar to an L1 sound in terms of formant frequencies and the position of articulatory organs the English vowel /iː/ can be expected to be completely substituted by the subjects' Bulgarian /i/.

The short vowel /ɪ/, on the other hand, is articulated as a considerably lower and retracted vowel with articulatory and acoustic characteristics indicating a more central position in the vowel space, compared to the peripheral position of its long equivalent. Having these considerably

different characteristics, it opens up two possibilities: it can either exhibit transfer from L1, in which case the subjects will have a considerable foreign accent, or it can be acquired as a new category.

Table 1. Formant frequencies of the English vowel /iː/ and the Bulgarian vowel /i/

formants	SSBE /iː/	Subjects' English /iː/	Subjects' Bulgarian /i/
F1	319 Hz	352 Hz	372 Hz
F2	2723 Hz	2533 Hz	2414 Hz

The data in Table 1 show that the formant frequencies of the English sound /iː/ produced by the subjects are more similar to those of their Bulgarian /i/ rather than to the formant frequencies of the English long sound. This means that there is minimal or no modification of their L1 sound /i/ so we can speak of a substitution of the English long front vowel by the L1 sound.

Another characteristic of the subjects' production of the long English front vowel which supports the conclusion that the L2 phoneme is substituted by the L1 sound is the absence of diphthongal pronunciation which gives a conspicuous auditory effect of L1 transfer.

The diphthongal pronunciation of the English /iː/ most clearly audible in final open syllables (Cruttenden 2008; Collins & Mees 2009) is considered an important characteristic of the sound. This diphthong is a glide from a somewhat central and lower position to the frontmost high position in the vowel space, and is commonly represented by the IPA symbol [ɪi]. The diphthongal quality of the vowel /iː/ is marked in the spectrogram by a gentle rise of F2 from the beginning of the vowel to approximately half of its duration and an almost unnoticeable gradual fall of F1 throughout the duration of the vowel.

Figure 1 represents the plotted graph of the native speakers' pronunciation of the English sound /iː/ and the subjects' production of the L1 sound and the English sound /iː/.[2]

Figure 1. Plotted F1:F2 graph of SSBE /iː/and the subjects' Bulgarian /i/ and English /iː/



Table 2 shows the mean F1 and F2 values of the short vowel /ɪ/ produced by the subjects.

Table 2. Formant frequencies of the English vowel /ɪ/and the Bulgarian vowel /i/

formant	SSBE /ɪ/	Subjects' English /ɪ/	Subjects' Bulgarian /i/
F1	432 Hz	443 Hz	372 Hz
F2	2296 Hz	2214 Hz	2414 Hz

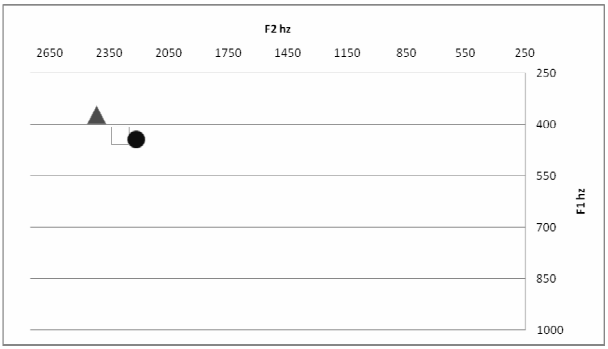
The data in Table 2 clearly indicate that the subjects have developed the L2 category with almost no transfer from L1. This supports the second prediction made at the beginning of the study and also confirms Flege's

[2] The same system of symbols is used in all F1:F2 graphs in the paper: subjects' Bulgarian vowels ▲, English vowels produced by the subjects ●, and English vowels produced by native speakers (SSBE) □.

model of phoneme acquisition and is in compliance with the results from Marković's study of the acquisition of high English vowels by native speakers of Serbian (Marković 2009). At this level of learning English, the subjects have obviously recognized the existence of the centralized, /ə/-like pronunciation of the short vowel /ɪ/.

In the graph in Figure 2, it can be seen that there is slight overlapping in the vowel spaces of the native speakers' short /ɪ/ and that of the same sound produced by the subjects.

Figure 2. Plotted F1:F2 graph of SSBE /ɪ/and the subjects' Bulgarian /i/ and English /ɪ/



4.2 Back vowels

The production of the pair of English vowels still traditionally referred to as 'back' poses a serious problem to Bulgarian learners of English due to the striking changes both /uː/ and /ʊ/ have undergone in contemporary English owing to considerable articulatory fronting (Cruttenden 2008; Collins & Mees 2009).

This tendency has been particularly noticeable in the pronunciation of /uː/ by younger generations. In acoustic terms, it is reflected in rather high values of F2, compared to the low values of the typical high back vowel [u]. In Bulgarian, on the other hand, /u/ is a fully back vowel for which the tongue is placed slightly below the closest position. The lips are energetically rounded and even pushed out. The short vowel /ʊ/ is a lower and centralized vowel compared to the typical back realization. However,

due to the process of fronting of the long vowel /uː/ in contemporary English, it has lower F2 values.

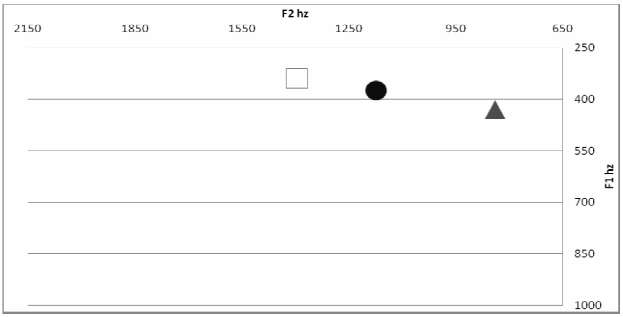
On the basis of this information, we can again make several predictions regarding the level of acquisition of these vowels. On the one hand, they are significantly different from the subjects' L1 vowel. Therefore, either a high level of L1 transfer and the presence of a strong foreign accent can be expected, or else a better acquisition owing to the different sound quality can be demonstrated (as in the case of /ɪ/, cf. Flege 1995). On the other hand, the task of acquiring the long vowel /uː/ in the subjects' interlanguage is further compounded by its fronting, as a result of which a vowel of particularly marked quality, rarely attested in the world's languages, is produced.

The mean values of the subjects' production of the English high back vowels confirm that their strategies of acquiring the back vowels are different from those observed in the acquisition of the front vowels. Table 3 shows the data on the subjects' production of the long English vowel /uː/.

Table 3. Formant frequencies for the English vowel /uː/ and the Bulgarian /u/

formant	SSBE /uː/	Subjects' English /uː/	Subjects' Bulgarian /u/
F1	339 Hz	374 Hz	430 Hz
F2	1396 Hz	1174 Hz	840 Hz

Figure 3. Plotted F1:F2 graph of SSBE /uː/and the subjects' Bulgarian /u/ and English /uː/



The results show that the subjects have modified their L1 sound in the direction of the L2 target sound but they have not completely achieved the native-like quality. These findings are easily observed in the F1:F2 graph (Fig. 3) which shows that the English vowel produced by the subjects is more or less half way between the L1 vowel /u/ and the English sound /uː/ but is closer to the English vowel sound.

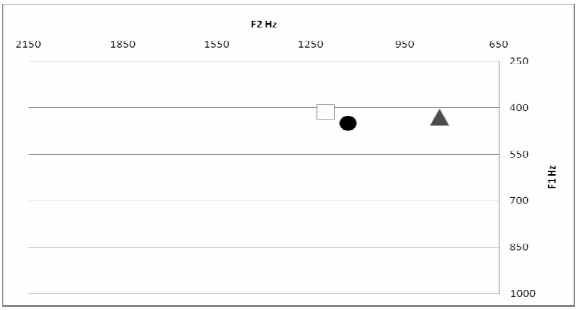
Another problem regarding the acquisition of the long vowel /uː/ again is the tendency towards diphthongization, with a glide of an [ʊu] type, widely spread among the speakers of the younger generation, with almost no lip rounding. For all of these auditory, acoustic, and articulatory characteristics, the subjects' production of this vowel gives the impression of a low level of acquisition but with an indisputable modification towards the target vowel.

The short vowel /ʊ/, which is, more or less the symmetric back equivalent of the front short vowel /ɪ/, seems to be better acquired by the subjects of the study. Table 4 provides the mean values of F1 and F2 frequencies.

Table 4. Formant frequencies for the English vowel /ʊ/ and the Bulgarian /u/

formant	SSBE /ʊ/	Subjects' English /ʊ/	Subjects' Bulgarian /u/
F1	414 Hz	450 Hz	430 Hz
F2	1203 Hz	1132 Hz	840 Hz

Figure 4. Plotted F1:F2 graph of SSBE /ʊ/ and the subjects' Bulgarian /u/ and English /ʊ/



As can be seen from the data, the subjects' formant frequencies have not completely reached the values of the native speakers' production but the close approximation to them is obvious. The auditory effect of the subjects' production of this sound on the examiner also confirms the findings that their pronunciation is similar to the L2 target vowel. The graph in Figure 4 also shows that the L2 vowel category is not completely acquired but is very closely approximated.

5. Discussion

The analysis of the data shows that the subjects exhibit different levels of acquisition of the four high English vowels, ranging from almost complete substitution by a similar L1 sound, through partial modification in the direction of the L2 target category to a high level of acquisition. The production of the high English vowels by the subjects of the study, who are proficient learners whose native language is Bulgarian, supports the claims of Flege's Speech Learning Model, namely that the more similar a foreign sound is to an L1 category, the more likely it is to be assimilated into an L1 category, and vice versa, the more different it is, the more likely it is to be earlier developed into an adequate L2 category because it is recognized by the learners of the foreign language as a new category.

Thus, of all the sounds studied here, the short vowels /ɪ/ and /ʊ/ demonstrate the highest level of acquisition. This may be because the subjects have recognised the presence of a central element in them and have successfully incorporated it in their production. The highly marked fronted /uː/ vowel of contemporary English has obviously triggered the process of new category formation in the learners' interlanguage and it can be said that the subjects of the study have achieved an acceptable level of acquisition of the sound but still have not fully acquired it. The vowel that exhibits the highest level of L1 transfer is the long vowel /iː/ probably because among the vowels studied here it is the one that is most similar to the sound in the subjects' native language.

6. Conclusion

This paper presents the results of a preliminary study of the production of the high English vowels by Bulgarian speakers of English, namely first year university students, and discusses the results of formant frequencies found during the spectral analysis of production of these vowels and the level of their acquisition by the subjects as well as assesses the degree of

L1 transfer in the subjects' pronunciation and the presence of 'foreign accent'. It should be noted that apart from the spectral characteristics there are various other reasons for vowels to sound foreign, such as their length, absence of pre-fortis clipping, different pre-and post-consonantal transitions, as well as prosodic characteristics which are not discussed in this study but can be subjects of separate studies. However, it is essential that when teaching pronunciation to foreign students we teach all these characteristics simultaneously at all levels of language learning. The results of the study show that at higher levels of learning, especially at tertiary level, it is important that we compare L1 and L2 phonological systems and provide empirical findings to raise the students' awareness of the sounds and their nature and help the students find their own ways through the difficult task of acquiring L2 phonology.

References

Best, C. 1995. A direct realist view of cross-language speech. In *Speech perception and linguistic experience,* edited by W. Strange, 171–204. Baltimore: York Press.

Boersma, P. & D. Weenink. 2005. Praat: doing phonetics by computer (Version 4.3.36) [Computer program]. Retrieved 2nd February 2012 from http://www.praat.org.

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

Celce-Murcia, M., D. Brinton & J. Goodwin, with B. Grinner. 2011. *Teaching pronunciation: A Course Book and Reference Guide.* Second edition. Cambridge: CUP.

Cenoz, J. M. & L. G. Lecumberri. 1999. The Acquisition of English Pronunciation: Learners' Views. *The International Journal of Applied Linguistics* 9: 3–18.

Collins, B. & I. Mees. 2009. *Practical Phonetics and Phonology. A Resource Book for Students.* Second edition. London: Routledge.

Cruttenden, A. 2008. *Gimson's Pronunciation of English.* Seventh edition. London: Hodder Education.

Danchev, A. 1990. On the contrastive phonology of the stressed vowels in English and Bulgarian. In *Papers and studies in contrastive linguistics,* Vol. 25, edited by J. Fisiak, 131–145. Poznan: Adam Mickiewicz University.

—. 1988. Segmental Phonology of the Bulgarian English Interlanguage(s). In *Error Analysis-Bulgarian Learners of English*, edited by A. Danchev, 156-175. Sofia: Narodna Prosveta State Publishing House.

Deterding, D. H. 1997. The formants of monophthong vowels in standard southern British English pronunciation. *Journal of the International Phonetic Association* 27: 47–55.

Despotova, V. 1978. Akusticen analiz na anglijskite fonemi /iː/ and /ɪ/, proizneseni ot balgari. *Contrastive linguistics* III/2: 28–36.

Dimitrova, S. 2004. *English Pronunciation for Bulgarians*. Sofia: Vezni 4.

Escudero, P. 2002. The Perception of English vowel contrasts: acoustic cue reliance in the development of new contrasts. In *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second Language Speech*, edited by A. James and J. Leather, 122–131. Klagenfurt: University of Klagenfurt.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In *Speech perception and linguistic experience: Issues in cross-language speech research*, edited by W. Strange, 233–277. Timonium, MD: York Press.

Flege, J. E., et al. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437–470.

Hawkins, S. & J. Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of theInternational Phonetic Association* 35 (2): 183–199.

Marković, M. 2009. Different Strategies in Acquiring L2 Vowels: The Production of High English Vowels by Native Speakers of Serbian. In *Ta(l)king English Phonetics Across Frontiers*, edited by B. Čubrović and T. Paunović, 3–18. Newcastle upon Tyne: Cambridge Scholars Publishing.

Minkoff, M. 1973. *An Introduction to English Phonetics*. Third edition. Sofia: "Naouka i izkoustvo" Publishing House.

Rauber, A., P. Escudero, R. Bion, & B. Baptista. 2005. The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. In *Proceedings of the Interspeech 2005 - Eurospeech*, 2913–2916. Lisbon: ISCA.

Stojanov, S. 1993. *Gramatika na balgarskja knizhoven ezik.* Sofia: Sofia University Publishing House "St. Kliment Ohridski".

Thomas, E. R. 2011. *Sociophonetics: An Introduction*. Basingstoke: Palgrave Macmillan.

ORIGINAL PRONUNCIATION: THE ACCENT OF SHAKESPEARE'S LONDON

ANDREJ BJELAKOVIĆ

Outline

The aim of this paper is to provide a brief sketch of Early Modern English (EME) pronunciation. Drawing from major works that deal with EME phonology, the paper lists all the main differences in pronunciation between the present-day English (PDE) and that of the 17th century. Each vowel phoneme is described separately, with the standard lexical sets of contemporary English used as the starting point. This was done in order for the paper to be more accessible to the potential lay reader, and to make it a generally useful reference tool for dealing with the history of English pronunciation. Also provided are examples in the form of phonemically transcribed verses taken from the works of William Shakespeare. One thing that is made clear by the paper is that EME is not undeserving of its name, since the sound system is phonologically rather similar, albeit phonetically different, from that of its present-day counterpart. Also evident is the fact that mainstream EME pronunciation shares many features with some regional accents of PDE, which is not surprising as in many ways it is their direct ancestor.

1. Introduction

Recent years have seen a resurgence of interest in the so-called Original Pronunciation (OP) [1], a reconstruction of the general accent used

[1] The term was pioneered by David Crystal in 2004, when he worked as the linguistic advisor on the Globe Theatre production of *Romeo and Juliet* in OP, the first such production since John Barton's 1952 staging of *Julius Caesar* in Cambridge; this was followed by *Troilus and Cressida* in 2005. Crystal continued

on the Shakespearean stage, and, more broadly, the accent current in Elizabethan and Jacobean London. The purpose of this paper is to provide a brief scholarly illustration of this Early Modern English (EME) pronunciation. For its phonological insight, it will draw on the major works from the past sixty years that focus on EME. For each sound, an example will be given in the form of a line taken from one of Shakespeare's works, presented both in modernised spelling and in phonemic transcription.

A common procedure in works of historical linguistics that deal with EME (e.g. Dobson 1957; Lass 1999) is to use the Middle English sound system as a starting point, and then chronologically trace the developments that led to the one found in EME. Going the other way around, we shall instead make the present-day English (PDE) a point of reference, and use the standard lexical sets proposed by John Wells, which are based on Received Pronunciation and General American[2]. This concept involves referring to a phoneme and the group of words that have that phoneme in the stressed syllable by a set keyword. So for example under 'FLEECE' I shall discuss what kind of a vowel was found during the Early Modern period in words that today have /i:/. Also included will be the most notable exceptions, i.e. words that switched from one set to another at some point during the intervening centuries. This will hopefully, among other things, make the paper more useful to speakers of PDE who seek, as it were, a phonetic map for the great authors of the past[3].

---

in succeeding years to popularize OP with his son, the actor Ben Crystal, in numerous talks and lectures around the world, most recently as a part of the Evolving English exhibition at the British Library. In November 2010 Paul Meier, a leading voice and dialect coach, staged a production of *A Midsummer Night's Dream* at the University of Kansas, the radio version of which was aired on Kansas Public Radio in April 2012. In November 2011 Robert Gander directed Hamlet in OP at the University of Nevada. For other OP-related events, see the website created by Crystal at http://originalpronunciation.com/.

[2] They were first used in Wells's *Accents of English* (1982), and have been subsequently widely adopted.

[3] It can be said that the general pronunciation described here was used both by Shakespeare and Milton, and also the one originally used for King James Bible.

2. Vowels

2.l. PRICE and MOUTH

By about 1500, the PRICE and MOUTH vowels had become diphthongs, stemming from the Middle English monophthongal /i:/ and /u:/ respectively (Barber 1997: 106; Lass 1999: 72). The diphthongs were narrow at first, /ɪi/ and /ʊu/, subsequently becoming wider, until the first part of the diphthong eventually reached the present fully open realisation (at least in RP and GA). The stage reached during the Elizabethan period involved both vowels either having a central onset (/əɪ/, /əʊ/), or having respectively front and back onsets (/eɪ/, /oʊ/ or /ɛɪ/, /ɔʊ/).[4]

Exceptions include words that had the MOUTH vowel before /r/, which had, in addition to the expected /əʊr/, another common possibility, /o:r/ (as in FORCE; see 2.6. below). This, along with **h**-dropping (see 3.1 below), makes possible the Shakespearean pun *hour – whore*, both being /o:r/ (Kökeritz 1953: 248).

Example:

Shall I compare thee to a summer's day?
Thou art more lovely and more temperate
ʃæl **əɪ** kəmˈpɛ:r ðɪ tʊ ə ˈsɤmərz dɛ:
ðəʊ a:rt mo:r ˈlɤvlɪ ənd mo:r ˈtempəˌrɛ:t

Sonnet XVIII, 1–2

2.2. GOAT

This lexical set came about through the coalescence of two distinct vowels in the late ME period. These were /ɔ:/ (as in ME *boat*, *oak*, *home*), and /ɔu/ (as in ME *know*, *grow*, *owe*). After the merger the resulting vowel became closer, yielding /o:/. As evidenced by numerous rhymes by Shakespeare, the merger was already largely complete by his time, although some conservative speakers still had a diphthong in the latter set (especially in the medial position) as late as the mid-17th century (Lass 1999: 93). Notable exceptions include the word *Rome*, which belonged to

[4] Lass (1999: 81–82) is the main proponent of the latter hypothesis; most other authors posit the former variants.

the GOOSE set (See Dobson (1957: 674–680) for the discussion of other such potential exceptions).

Example:

The evil that men do lives after them;
The good is oft interred with their bones
ðɪ iːvɪl ðət men duː lɪvz ˈæːftər ðəm
ðə gʊd ɪz ɒːft ɪnˈterɪd wɪθ ðɛːr **boːnz**

Julius Caesar, Act III, Scene 2

2.3. GOOSE

The present-day GOOSE set comprises:

a) words that had the ME /oː/ (although many of them shifted to FOOT and STRUT, e.g. *foot*, *blood*), as in *food*, *tooth*. This vowel started to close in the 14th and continued to do so throughout the 15th century. By the early 16th century the vowel had become a fully close /uː/ (Lass 1999: 80), and indeed this is what we find in Elizabethan pronunciation. It is worth keeping in mind that this was a fully back, rounded vowel unlike the modern day central realisation (i.e. more like the sound in Serbian *ruka* or German *gut* than the one in contemporary English *goose*);

b) words that in ME had diphthongs /iu/ (*new*, *blue*, *truth*) and /ɛu/ (*dew*, *shrewd*, *beauty*, *few*). These two vowels were either merging or had already merged to /iu/ during Shakespeare's lifetime. It is safe to assume that by 1700 both sets ended up having /juː/, and in the centuries that followed /j/ disappeared from some of these words (Barber 1997: 133; Lass 1999: 98–99).

To sum up, in the period examined here *tooth* was /tuːθ/, while *truth* was /triuθ/.

Example:

If music be the food of love, play on
ɪf ˈ**miuzɪk** biː ðə **fuːd** əv lɤv pleː ɒn

Twelfth Night, Act I, Scene 1

2.4. FLEECE

The lexical set containing the words that have /iː/ today has had somewhat of a complex history, coming about through a series of mergers of earlier lexical sets. We can divide these into two groups:

a) words with *ee* or *ie* in spelling (*fleece*, *piece*), i.e. those that had /e:/ in ME. By about 1500 this vowel shifted to /i:/, which is indeed what we find in PDE.

b) words spelt with *ea* or *ei*, and some spelt *e*C*e* (e.g. *sea*, *conceit*, *complete*), which had /ɛ:/ in ME. These words had two possible pronunciations current in Elizabethan English. One was /ɛ:/~/e:/; the other was /i:/, which meant a merger with the words in subset a). The latter variant seems to have been at first a popular London feature and a general Northern variant, only later spreading to the rest of England and to more formal types of speech, finally becoming standard, and the only surviving variant[5]. Since both existed in Shakespeare's time, we find him using both, depending which served best his rhyming purposes: *these* and *seas* (/ðiːz/, /siːz/), but also *sea* and *play* (/sɛː/, /plɛː/ or possibly /seː/, /pleː/, see 2.5. below under FACE). It is probably best to use the /ɛ:/~/e:/[6] pronunciation in the subset b) unless the rhyme requires /i:/ (Kökeritz 1953: 194–200; Barber 1997: 107–108; Lass 1999: 95–98).

Notable exceptions include *shield*, *field,* and *yield*, which belonged to KIT (Kökeritz 1953: 192).

Example:

So long as men can breathe or eyes can see,
So long lives this, and this gives life to thee.
so: lɒŋ əz men kən **brɛ:ð** ər əɪz kən **si:**
so: lɒŋ lɪvz ðɪs ən ðɪs gɪvz ləɪf tʊ **ði:**
Sonnet XVIII, 13–14

2.5. FACE

As with most other sets examined so far, this present day set comprises two groups of words that used to have different vowels in their stressed syllables:

a) Words with *a*C*e* in spelling (*name*, *age*, *bathe*), which had /a:/ in ME, a vowel that through raising later became /æ:/, and finally /ɛ:/.

[5] A handful of words however ended up joining the FACE set, and thus do not have /i:/ today. These are *break*, *great*, *yea*, and *steak* (Lass 1999: 96).

[6] Note that if we posit /ɛ:/ here instead of /e:/ it means that b) words have the same vowel as the FACE set (see below). For simplicity's sake, this will be the course pursued in this primer.

b) Words with *ai*, *ei*, *ay*, or *ey* in spelling (*wait*, *veil*, *day*, *they*), which had /ai/ in ME. By the late Middle English period, the first element of this diphthong became raised, yielding /æi/, which was followed by subsequent monophthongisation (Barber 1997: 107, 114).

In the century and a half between 1500 and 1650 the vowels in these two groups of words, plus words from the FLEECE subset b) had potentially a very complicated history (for details see Lass 1999: 91–94). For purposes of this primer we shall assume, as is done by Kökeritz (1953: 173), that a) and b) had merged by Shakespeare's time, and had /ɛ:/ (as was certainly the case with at least some speakers).

Example:

Let me not think on't—Frailty, thy name is woman!
let mi: nɒt θɪŋk ɒnt **ˈfrɛ:ltɪ** ðɪ **nɛ:m** ɪz ˈwʊmən
Hamlet, Act I, Scene 2

2.6. CHOICE

Words that in present-day English have /ɔɪ/ in ME had either /ɔi/ or /ui/. Those that were pronounced with /ui/ had the first part of the diphthong unrounded, much like ME /u/, (see 2.14. below), which led to the merger of ME /ui/ and ME /i:/ under /əi/ (see 2.1. above). As is often the case, the phonological process was not quite as neat and some words for some speakers transferred from ME /ɔi/ set to the merged /əi/ set. In other words, some present day /ɔɪ/ words had /ɔɪ/ (or perhaps /ɒɪ/) in EME, and some had /əi/. The shift back did not occur until the late 18th century, after which time we encounter the CHOICE set pretty much as it is today. Shakespeare himself rhymed *die* and *joy* and *annoy, right* and *exploit*, *bile* and *boil, line* and *loin, vice* and *voice*. Words that seem never to have merged, i.e. those that retained /ɔɪ/ throughout include *boy*, *toy*, *choice*, *noise* and some others (Kökeritz 1953: 216–221; Lass 1999: 102–103).

Example:

And, nuzzling in his flank, the loving swine
Sheathed unaware the tusk in his soft groin

ənd nɤzlɪŋ ɪn ɪz flæŋk ðə lɤvɪŋ **swəɪn**
ʃɛ:ðd ɤnəˈwɛ:r ðə tɤsk ɪn ɪz sɒ:ft **grəɪn**
Venus and Adonis, 1115–1116

2.7. PALM and BATH

The PALM lexical set, words that today have /ɑ:/ in most accents, did not exist as such in EME. Most words of foreign origin that belong to this set in PDE had not yet entered English. The remaining others, words of native origin like *calm*, *balm*, *palm*, *almond*, *father*, *half* had /a:/ or /æ:/, like in BATH.

BATH consists of words that today have /æ/ in GA, /a/ in the North and Midlands of England, /ɑ:/ in RP and most of the accents of the South, and some sort of an intermediate form in most other accents of English (Wells 1982: 133–135, 232–233). In other words, the vowel in this lexical set lengthened and retracted at some point in southern English accents. By Shakespeare's time the former had occurred, but the latter most likely had not started yet. This means that in *bath*, *staff*, *castle* etc. we get /æ:/~/a:/ – the same vowel as in TRAP, only long. Since the words belonging to this set that have a nasal plus obstruent following the vowel (*dance*, *chance*, *lance*, *plant*, *chant* etc.) used to have /au/ in ME, many of them at this point had doublets with /ɒ:/, the usual reflex of ME /au/, and thus optionally belonged to the THOUGHT set (see 2.8. below). Indeed, *haunt, vaunt* and *launch* permanently shifted to THOUGHT, and in time accordingly changed their spelling to 'au' (Lass 1999: 103–107).

Example:

Royal Lear,
Whom I have ever honour'd as my king,
Lov'd as my father, as my master follow'd,

lʏvd əz məɪ **ˈfa:ðər** əz məɪ **ˈma:stər** ˈfɒlo:d

King Lear, Act I, Scene 1

2.8. THOUGHT and CLOTH

The present day THOUGHT set had been largely formed by the Elizabethan era. The vowel used in the words belonging to this set had come about via ME /au/ which had monophthongised by Shakespeare's time, and yielded a fully open, lightly rounded back vowel, /ɒ:/ (although there is evidence that a diphthong of the /ɒu/ type was still to be found in some conservative speech) (Lass 1999: 94–95). The same vowel was found in the CLOTH set, comprising words that had short ME /o/ which lengthened before voiceless fricatives (e.g. *cloth*, *off*, *lost* etc.). In present day AmE these words still have the same vowel as in THOUGHT; however

in British accents the process was reversed and the words in question went back to the LOT set (Wells 1982: 136, 234).

Example:

For in that sleep of death what dreams may come
When we have shuffled off this mortal coil,
Must give us pause [...]
hwen wi: əv ʃɤfəld **ɒ:f** ðɪs mɒ:rtəl kəɪl
məst gɪv əs **pɒ:z**

Hamlet, Act III, Scene 1

2.9. LOT

This set, containing words that had ME /o/ (except those that underwent lengthening; see 2.8. above), was to a large extent the same as it is today. By the late 16th century the vowel had lowered, yielding /ɔ/ ~ /ɒ/. Worth mentioning is the realisation used by some, though deemed an affectation by others, employed as a "well-known foppish stereotype in restoration and later drama", which included a completely unrounded vowel. As a consequence we see such spellings as Queen Elizabeth's *stap* for *stop*, Shakespeare's rhymes such as *dally* and *folly (*from *The Rape of Lucretia*) or puns such as *tropically* spelt as *trapically* (in *Hamlet*) (Lass 1999: 86–87). This was especially common before /r/ (see 2.16. below).

As regards the words which had 'a' in the spelling and had /a/ before /w/ in ME, (*wand*, *swan*, *want*, *watch* etc.), they still had an unrounded vowel in Shakespeare's time (he rhymes *wand* and *hand* for example), and thus belonged to the TRAP set, though they probably never had the raised /æ/ (see 2.10. below). The rounding started in the 17th century and became general in mid-18th (Kökeritz 1953: 171–172; Barber 1997: 122; Lass 1999: 86).

Example:

But soft, what light through yonder window breaks?
bət sɒ:ft hwat ləɪt θru: ˈ**jɒndər** ˈwɪndo: brɛ:ks

Romeo and Juliet, Act II, Scene 2

2.10. TRAP

The ME /a/ seems to have started raising and reached the modern-like /æ/ (this clearly never happened in the Midlands and North of England) precisely in the period under examination here. According to some

scholars, this had occurred before the time of Shakespeare's birth, while according to others the process took place in the first half of the 17th century (Kökeritz 1953: 162; Lass 1999: 85). In any case, words such as *watch*, *what* etc. (see 2.9. above) had the unraised allophone [a], or even the retracted [ɑ]. Other words had doublets belonging to the DRESS set (*catch*, *gather*, *January*, *jasmine* etc.) (Dobson 1957: 564; Cercignani 1981: 100; Lass 1999: 85).

Example:

That defunctive music can,
Be the death**-**divining swan
ðæt dɪˈfʏŋktɪv ˈmiuzik **kan**
bi: ðə deθ dɪˈvəɪnɪŋ **swan**
The Phoenix and the Turtle, 14–15

2.11. DRESS

Similar to other short vowels, the vowel in the Elizabethan DRESS set was much like the modern one. The ME /e/ had become somewhat less close, yielding an [ɛ]**-**like vowel. However, the distribution of words between DRESS and KIT sets fluctuated. Thus, many words that today belong to the former were, at least optionally, a part of the latter. These included words like *devil* (commonly spelt *divel*), *yes*, *yesterday*, *yet*, *get*, *together*, *well*, *red*, *engine* etc. (Kökeritz 1953: 186–188). Also, *many* and *any* belonged to the TRAP set, as they still do in Scotland and Ireland.

2.12. KIT

The vowel in the KIT set became the lax /ɪ/ in the ME period[7], and stayed much the same to the present day. As with DRESS, there is some slight variation when it comes to the words that belonged to this set (e.g. *spirit* and *mirror* were often pronounced with /e/, and *give* with /i:/) (Kökeritz 1953: 212–214; Cercignani 1981: 50–58).

[7] Although according to Lass (1999: 88) this did not occur until the second half of the 17th century.

2.13. FOOT

As with KIT, the vowel in this set had become by Shakespeare's time the lax /ʊ/ we know today. The distribution of words, however, between this set and the GOOSE and STRUT sets fluctuated greatly during this period. Of the present day FOOT words that, at least optionally, had different vowels there are *foot*, *wood*, *should*, *would*, *good*, *look* etc. with /u:/, and *wolf*, *bull*, *bullet*, *full*, *pull* which belonged to STRUT (see 2.14. below) (Kökeritz 1953: 236–237, 242; Lass 1999: 88–91).

2.14. STRUT

By Shakespeare's time, the FOOT–STRUT split had already been completed, at least in London. Though unrounded, the vowel in these words had centuries to go before becoming fully open. In the period under examination the vowel was found somewhere in the vowel space triangle formed by [ɯ], [ʌ], and [ə] (Kökeritz 1953: 240; Lass 1999: 90). Crystal prefers to transcribe it as /ɤ/, a practise continued here. Of the distribution discrepancies between PDE and EME, worth mentioning is that the words *one*, *none* and *done* were most commonly pronounced with /o:/ (like in GOAT), and that *does* and *doth* still had possible long vowel realisations (with /u:/ as in GOOSE). Also, *tongue*, *young*, *wrung* etc. belonged to LOT (as they still do in the English Midlands), and *come* and *love* seem to have had archaic optional realisations with /u:/ (Kökeritz 1953: 243–244; Dobson 1957: 508–510; Barber 1997: 136).

Example:

Golden lads and girls all must,
As chimney-sweepers, come to dust.
ˈgo:ldən lædz ən gəɪlz ɒ:l **mɤst**,
əz ˈtʃɪmnɪ ˈswi:pəɪz **kɤm** tə **dɤst**.

Cymbeline, Act IV, Scene 2

2.15. Happy

This weak vowel could in Shakespeare's time be pronounced either as /ɪ/, or as /əɪ/ as in PRICE (this seems to have been especially common in the *–ly* suffix). Hence the rhymes such as *eye*: *chastity* etc. (Kökeritz 1953: 219–220; Dobson 1957: 842–845).

2.16. Rhotic sets (Vowels before /r/)

Three important phonological changes had not yet occurred in Shakespeare's time: the loss of non-prevocalic /r/ (see 3.6. below), pre-/r/ breaking and pre-schwa laxing (Wells 1982: 213–222). These three would eventually turn /bi:r/ into /bɪə/. This means that words in lexical sets CURE, FORCE, NORTH, NEAR, SQUARE and START generally speaking all had /r/ preceded by the same tense monophthong found respectively in GOOSE, GOAT, THOUGHT, FLEECE, FACE and PALM (i.e. /u:/, /o:/, /ɒ:/, /i:/, /ɛ:/, /a:/) (Lass 1999: 108–111).

Seeing as most accents of PDE do not distinguish FORCE from NORTH (many Scottish and West Indian accents being a notable exception), it would be worth indicating which common words belonged to which set, although, as with other sets, there was some fluctuation present.

FORCE /o:r/: *ore*, *before*, *more*, *store*, *floor*, *door*, *fort*, *sport*, *pork*, *forth*, *ford*, *sword*, *hoarse*, *coarse*, *board*, *hoard*, *court*, *course*, *source* etc. Plus the words that had the MOUTH vowel before /r/, in addition to /əur/ had doublets with /o:r/ (*power*, *hour*, *flower*, *our* etc.)

NORTH /ɒ:r/: *or*, *for*, *George*, *short*, *cork*, *fork*, *stork*, *accord*, *lord*, *storm*, *corn*, *horn*, *corpse*, *horse* etc. (Wells 1982: 159–162). This does not include the words spelt with 'ar' that today belong to NORTH (*war*, *reward*, *warm*, *dwarf*), which, as mentioned above under LOT, still had an unrounded vowel, and belonged to START (Kökeritz 1953: 172; Lass 1999: 86).

Example:

Then should the warlike Harry, like himself,
Assume the port of Mars; and at his heels,
Leash'd in like hounds, should famine, sword and fire
ðen ʃud ðə **wa:r**ləık hærı ləık ımself
əsju:m ðə **pɒ:rt** əv ma:rz ənd æt ız hi:lz
lɛ:ʃt ın ləık həʊndz ʃud fæmın **swo:rd** ənd fəır
Henry V, Prologue

As regards the long front vowels before /r/, two major discrepancies between PDE and EME must be mentioned. The first is to do with the distribution of words between the sets NEAR and SQUARE: today words *beard*, *dear*, *ear*, *fear*, *gear* belong to the former; in Shakespeare's time they were in the latter, and accordingly had /ɛ:r/ (compare 'ea' words above under FLEECE) (Kökeritz 1953: 206–209).

Example:

Friends, Romans, countrymen, lend me your ears;
frendz ˈruːmənz ˈkʏntrɪmen lend mɪ juːr **ɛːrz**
Julius Caesar, Act III, Scene 2

The second involves the START set and part of the PDE NURSE set that has ‘er’ in modern spelling (*certain*, *heard*, *servant*, *person*, *mercy* etc.). These slowly shifted from the former to the latter set in the centuries between 1500 and 1800. In Shakespeare’s time the majority of them still had the same vowel as in START, /aː/ (hence the rhymes *convert* and *art*, *desert* and *impart* etc.) (Kökeritz 1953: 250–251; Dobson 1957: 560–561; Barber 1997: 118).

The PDE NURSE set came about through merger of words that had ME /er/, /ir/ and /ur/ (as in *earth*, *bird*, and *curse*). By the Elizabethan era the merger had definitely begun; specifically /ir/ and /ur/ seem to have coalesced, yielding /ər/, with /er/ still separate in most speech. In summary, some PDE NURSE words had /ər/, some /er/, and some /a(ː)r/ (Kökeritz 1953: 252–254; Dobson 1957: 759; Lass 1999: 112–113)

Example:

The quality of mercy is not strain'd
ðə ˈkwalɪtəɪ əv ˈ**marsəɪ** ɪz nɒt streːnd
The Merchant of Venice, Act IV, Scene 1

If it were done when 'tis done, then 'twere well
It were done quickly [...]
ɪf ɪt **wer** dʏn hwen tɪz dʏn ðen **twer** well
ɪt **were** dʏn ˈkwɪkləɪ
Macbeth, Act I, Scene 7

The private wound is deepest. O time most accursed,
’Mongst all foes that a friend should be the worst!
ðə ˈprəɪvɪt wəʊnd ɪz ˈdiːpɪst oː təɪm mːost əˈ**kərst**
mɒŋst ɒːl foːz ðət ə frend ʃʊd biː ðə **wərst**
The Two Gentlemen of Verona, Act V, Scene 4

3. Consonants

3.1. /h/

Word-initial /h/ in stressed syllables was much more prone to deletion in Shakespeare's time than it is today. In fact, it was on its way out, and was only later restored. The stigmatisation of /h/-dropping does not appear until the 1750s, and becomes widespread only in the 19th century (Kökeritz 1953: 307–309; Lass 1999: 118).

3.2. /ŋ/

As with /h/-dropping, the [ŋ] > [n] change at the end of gerunds, present participles and some other words was widespread, and not just confined to informal/stigmatised speech. For example, the famous Elizabethan, Philip Henslowe, wrote *ten shellens* for 'ten shillings', and the Queen herself wrote *besichen* for 'beseeching'. The modern pattern emerges in the 19th century (for example, Byron still rhymes *children* and *bewildering* etc.) (Kökeritz 1953: 313–314; Lass 1999: 119–120).

3.3. /ʃ/, /ʒ/, /tʃ/ and/dʒ/

In the period written about here consonants in clusters /sj/, /zj/, /tj/ and /dj/ found in unstressed syllables began to undergo coalescence and yield /ʃ/, /ʒ/, /tʃ/ and/dʒ/ (as in *malicious*, *vision*, *Christian*, *soldier* etc.). The word endings in question could also be made disyllabic, so for musician there was /mjuˈzɪsjən/, /mjuˈzɪʃən/ or /mjuˈzɪsɪən/. The exception are the endings *-ture*, *-dure*, *-sure* which had plain /t/, /d/, /s/ or /z/ (e.g. *nature* /ˈnɛːtər/, *censure* /ˈsensər/). The process was extended to initial stressed syllables when it came to /sju:/ (which yielded /ʃ/ in *sure*, *sugar*, but also in *sue* and *suitor*, which later went back to /sj/ (Lass 1999: 121–122).

3.4. wh-

The consonant cluster /hw/ had not yet simplified by the 17th century, so pairs such as *witch* and *which*, and *whether* and *weather* were not homophones (the majority of present day Scottish and Irish accents still resist the merger) (Wells 1982: 228–229; Lass 1999: 122).

3.5. /kn/ and /gn/

Whereas many other ME initial consonant clusters had already simplified by Elizabethan times, and /hw/ as mentioned above firmly resisted such fate, /kn/ as in *knight* and /gn/ as in *gnaw* were undergoing simplification in the period examined here. The first step was a merger of the two under /kn/. That and the further simplification /n̥n/ ~ /hn/ seem to have been the most frequent realisations in Shakespeare's time (Barber 1997:126; Lass 1999: 123).

3.6. /r/

All accents of English in Shakespeare's time were fully rhotic. However, the non-prevocalic /r/ had already begun to weaken. Thus we find two major allophones of this phoneme: the initial and intervocalic /r/, which was either a trill [r] or a tap [ɾ], and the preconsonantal and prepausal /r/, which was probably an approximant (either alveolar [ɹ] or retroflex [ɻ]) or even a fricative [ɹ̝] (Kökeritz 1953: 314–315; Lass 1999: 114–115).

4. Word stress

The history of stress placement in English words is, broadly speaking, a history of interplay and adjustment between the old Germanic stress rules and the newer Romance stress rules (Lass 1999: 125). As one would expect, in EME many words had different stress placement than the one found in PDE. This final section will provide a short list of such words, largely adapted from Kökeritz (1953: 392–398):

ˈabbreviation, ˈabsolute, ˈacademy, ˈacceptable, ˈaccessory, ˈaccommodate, ˈadjacent, adverˈtize, alˈcove, ˈallegorical, ˈantique, asˈpect, auˈthorize, chaˈracter, colˈleague, comˈmerce, ˈcomplacency, ˈcomplete, conˈfine (n.), conˈtents, conˈtract (n.), conˈtrary, ˈcontribute, ˈconvenient, ˈcorrosive, ˈcorruptible, ˈdefective, deˈmonstrate, ˈdetestable, ˈdistribute, ˈdivert, ˈexcusable, eˈxile (n.), ˈextreme, flaˈgon, floˈrin, galˈlantly, ˈhorizon, ilˈlustrate, inˈstinct, leˈgate, ˈobscure, ˈobservance, paraˈmount, paˈrent, ˈperfect (vb.), ˈperspective, porˈtent, ˈproclamation, proˈfile, proˈmulgate, reˈcord (n.), reˈtinue, reˈvenue, siˈnister, stubˈborne, ˈsuggestion,ˈsupportable, ˈsupreme, temˈporize, triˈumph (vb.), ˈunawares, yesterˈday,

Notably, trisyllabic words that had stress on the first syllable often had secondary stress on the final syllable, which consequently had an unreduced vowel. This makes possible rhymes such as *date* and *temperate* (ˈtempəˌrɛ:t), *age* and *pilgrimage* (ˈpɪlgɪrmˌɛ:dʒ), *advance* and *ignorance* (ˈɪgnərˌæ:ns), *day* and *Asia* (ˈɛ:sɪˌɛ:) (Barber 1997: 133–134).

5. Conclusion

Outlining the general structure of Early Modern English phonology in this primer we have tried to strike a balance between providing a too narrow and absolute view of the EME sound system, and delving too deep into some of the controversies still present among the historical linguists[8]. Namely, depending on whether one chooses to attach greater importance to the accounts of the orthoepists and other commentators of language issues (the EME period being the first one characterized by the presence of such writers), or to other kinds of evidence such as spelling and rhyme or comparative evidence, the accent used on the Elizabethan stage can be more or less modern. In other words, there is a window of about 150 years, spanning from the late 16th century to the very last decades of the 17th century, during which it is difficult to say with great certainty how advanced were many if not most of the changes outlined above.

Regardless, even if the general accent described here would have been more characteristic of Alexander Pope's rather than William Shakespeare's London, the fact remains that the pronunciation most likely used on the Elizabethan stage was a far cry from the Middle English one we associate with, say, Chaucer. Although phonetically different, it was phonologically remarkably similar to present-day English. The similarity is all the more apparent if one compares this reconstructed accent of the past to the pronunciation still found in many rural accents of Great Britain and Ireland.

[8] For an example of a chapter in one such long-standing debate see Minkova and Stockwell (1990); a more detailed discussion can be found in Lass (1999).

References

Barber, C. 1997. *Early Modern English*. Second edition. Edinburgh: Edinburgh UP.

Cercignani, F. 1981. *Shakespeare's Works and Elizabethan Pronunciation*. Oxford: Clarendon Press.

Crystal, D. 2005. *Pronouncing Shakespeare*. Cambridge: CUP.

Dobson, E. J. 1957. *English Pronunciation 1500–1700*. Oxford: Clarendon Press.

Kökeritz, H. 1953. *Shakespeare's Pronunciation*. New Haven: Yale UP.

Lass, R. 1999. Phonology and Morphology. In *The Cambridge History of the English Language, Vol. 3: 1476–1776*, edited by R. Lass, 56–186. Cambridge: CUP.

Minkova, D. & R. Stockwell. 1990. Early Modern English Vowels: More O'Lass. *Diachronica* VII/2: 199–221.

Shakespeare, W. 2005. *The Oxford Shakespeare: The Complete Works*. Second edition. Edited by J. Jowett, W. Montgomery, G. Taylor and S. Wells. Oxford: Clarendon Press.

Wells, J. C. 1982. *Accents of English*. Cambridge: CUP.

PART II:

SUPRASEGMENTALS AND BEYOND

PITCH ALIGNMENT IN WELSH ENGLISH: THE CASE OF RISING TONES IN GWYNEDD

STEFANO QUAINO

Outline

Welsh English and Celtic English in general are known for their preference for rising tones in declaratives. Gimson (2008), Cruttenden (1995), Tench (1990), Walters (1999) have thoroughly analysed a number of different varieties of British English, concluding that rising movements are extremely common in Ireland, Scotland and Wales. Gimson (2008: 289) states that "rises are more frequent on declaratives than in RP and [*sic*] may typically be the most frequent tone on declaratives". Cruttenden even considers the case of Liverpool English, stating that there is "an increase in the number of rises compared with RP", the cause of which could be related to "influx of people from Scotland and Ireland", so in "most of the areas involved [in the greater use of rises] there is a strong Celtic influence' (1995: 155).

However, stating that Welsh English has a preference for rising tones is not alone sufficient to explain its main prosodic characteristics. The aim of this paper is to analyse rising tones in Gwynedd English via PRAAT and underline their distinguishing features.

1. Introduction

Welsh English (WE) is a general term to describe the Englishes spoken in Wales which differ from RP and other British accents and varieties. Some of the most characteristic features of WE are its intonation and prosodic patterns, which have been defined by Walters (1999) as tuneful and "sing-song". As already mentioned, Celtic Englishes have a general preference for rising tones and this has been noticed also in WE.

The present research aims to be a continuation of Walters' work on Rhondda Valley English (RVE) and will focus on the English spoken in

Gwynedd (GE): this county in North Wales is known as the stronghold of the Welsh language, being the first language of most inhabitants. For this analysis, several excerpts from the interviews taken for the Survey of Anglo-Welsh Dialects have been collected and then processed through PRAAT.

It will also be interesting to compare the present results with those offered by Walters. A perfect comparison is not possible, since the methods used are not completely similar; in addition, two researchers do not necessarily arrive at the same result, even when using identical methods. However, comparing GE and RVE will underline the main similarities of the two varieties and thereby come closer to a prosodic characterization of a Welsh English accent in general.

2. Informants, alignment and methods of analysis

The present research is based on a number of excerpts from the Survey of Anglo-Welsh Dialects (SAWD). The SAWD was a project conceived by David Parry, a dialectologist at Swansea University, in the late 1960s. After preparing a set of questions, Parry and his colleagues (starting from the 1970s) began to visit many villages in Wales and interview local people (aged between 60 and 80); the project continued till the 1990s, when a comprehensive volume with much information about the phonetics, phonology, syntax, morphology and vocabulary of Welsh English was able to be published. The interviews which have been used for the present research were prepared between the late 1970s and the early 1980s.

The present prosodic analysis has been initially carried out via a modified version of ToBI, a method which has also been used by Walters in his research on Rhondda Valley English (1999). The attention will be focused on contours found at pitch accents, which are composed of two distinguishable movements:

- The first contour point is the movement towards the stressed syllable.
- The second contour point is the movement from the stressed syllable.

Intonation analysis will be carried out via the measuring and classification of the contour points of utterances: H will identify a contour point which is higher than the previous one, which means that the pitch has risen; L will be used to describe a contour point lower than the

previous one. If the pitch does not present relevant upwards or downwards movements, the specific contour point will be marked with a 0. As a result, H*+H describes a pitch accent whose both first and second contour point are rising; L*+H, instead, identifies a pitch accent which is formed by an initial fall and a subsequent rise.

However, if one wants to present a more detailed analysis of the language, it is not sufficient to determine whether the pitch movement is upward or downward, nor is it by stating the numerical value of the contour points; ToBI analysis alone cannot present the real distinguishing prosodic features of a language. Since pitch movements are found in every language, one should wonder whether they share the same characteristics or present differences: for example, matters of extreme interest are those related to the starting point of the rise and the position of the pitch peak. Alignment is the theoretical concept which comprises this kind of analysis. Although many aspects of alignment remain obscure, it can nonetheless be stated that alignment is "highly lawful and can be systematically influenced by a range of phonetic and phonological effects" (Ladd 2008: 169). Alignment is strictly connected to the realisation of the tones and is "a phonetic property, namely the relative timing of events in the F0 contour and events in the segmental strings" (Ladd 2008: 179).

When analysing alignment, two features constitute the main focus of interest, namely the beginning of the rise and the position of the pitch peak. It is necessary to determine the exact position of the starting point of the upward movement, since certain languages show a preference for early rises, while others prefer late rises.

After establishing the starting point of the upward movement, attention is then shifted to the pitch peak. The H tone is an instance of rising intonation which carries the pitch from a lower to a higher point, so that during its pronunciation the F0 will be characterized by an upward movement. At one specific point within the syllable, the pitch will reach its peak: it is interesting to determine whether the peak has occurred towards the onset of the syllable (beginning), the midpoint, or the end.

The present paper deals with the pitch alignment in Gwynedd English: the interest will focus on those contours which carry either a rising or a level movement, regardless if it is towards or from the stressed syllable. Although rises and level tones are not completely similar, they are both non-falling movements.

In addition, pitch accents should also be given the Most Prominent Semantic Element (MPSE) status or found at the very end of an intonational phrase, so that they carry a terminal tone. The MPSE

identifies the word which has been given the major focus by the speaker; talking is a way of communicating information with others, with speakers always having a specific goal in their mind. They want to share a message with others, and intonation is a major tool for this orientation. Utterances have been analysed as responses to previous questions and if no obvious question/answer structure is present, two or three previous statements in the communication have been taken into consideration.

The MPSE could be compared to the nucleus, but I prefer to distinguish the two terms. The nucleus, also known as tonic syllable, is the "most prominent, or salient, of the stressed syllables in a given unit' (Tench 1996: 53). Since the nucleus is a strongly accented syllable, its identification is enabled via tone and intensity values: the tonic syllable is usually marked by the highest pitch or the greatest pitch movement throughout the unit; along with this, it is usually introduced by a crescendo and followed by a decrescendo. The analysis of rising tones will be carried by considering a series of questions, such as:

- Is there any delay in the alignment of the stressed syllable?
- Is the intensity peak aligned at the beginning of the syllable or with a delay?
- Is the upward movement towards the stressed syllable rapid or not?
- What occurs to the pitch after the stress?
- Where is the pitch peak aligned? Immediately after the stress or is it delayed?

Only through the detailed analysis of these key elements will it be possible to present the distinguishing features of Gwynedd and thereby come closer to a prosodic characterization of a Welsh English accent.

List of informants

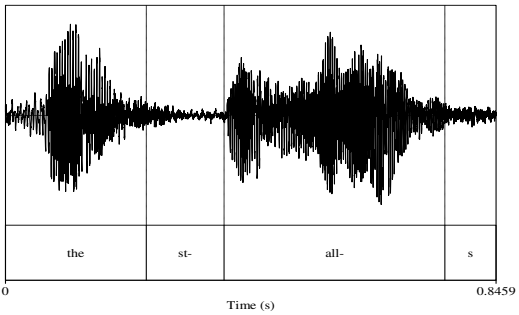
- Informant #1 (Trefor, Female)
- Informant #2 (Ynys, Female)
- Informant #3 (Dolgellau, Female)
- Informant #4 (Botwnnog, Male)
- Informant #5 (Botwnnog, Male)
- Informant #6 (Dolgarrog, Male)
- Informant #7 (Gyffin, Female)
- Informant #8 (Trefor, Male)

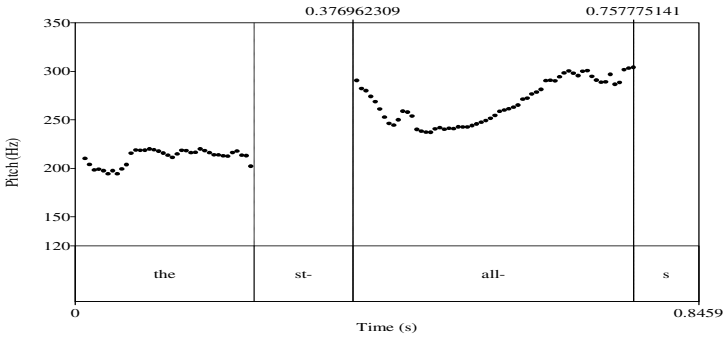
3. Rise

3.1. Rise on the second contour point (H*+H; 0*+H)

The initial analysis will focus on those pitch accents which present a rise (or a non-falling movement, i.e. a level tone) in the second contour point, such as H*+H, H*+0, 0*+H and L*+H. The analysis starts with two utterances taken from an interview with a lady from Trefor.

Example #1 (Informant #1)





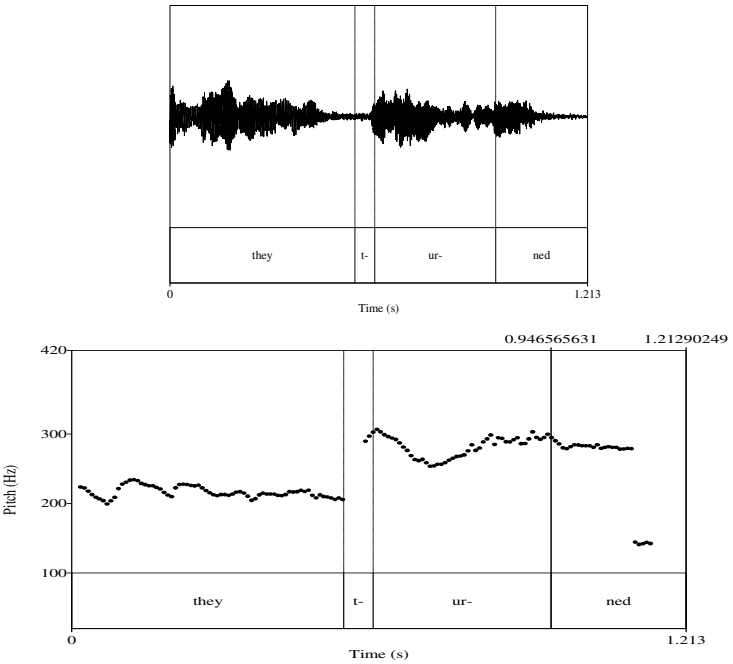
Example #1 focuses on the pronunciation of "stalls". As explained in the introduction, there are two elements to consider in a pitch accent, namely the positioning of the first intensity peak (i.e. alignment of the first contour point) and the beginning of the actual rise.

As shown by the graph, the pitch rises after "*the*", so that the first contour point is obviously marked with an H*. It appears that it is aligned

with very limited delay, since the first intensity peak occurs less than 30 ms. after the consonant clusters has been released into the vowel [□:]. The analysis of the pitch line at "*stalls*", instead, reveals that the rising movement is not swift, but presents a 'sagging' effect, i.e. a curve. As shown by the graph, the actual rising movement (i.e. the second contour point, the H) does not start before 130 ms. into the syllable, being preceded by a short fall.

This feature is extremely important: similar research on other varieties of Welsh English, such as Ceredigion English, has revealed that the alignment of the first contour point can present a much bigger delay (even 80 ms.). For this reason it is crucial to consider other examples and determine the positioning of the intensity peak.

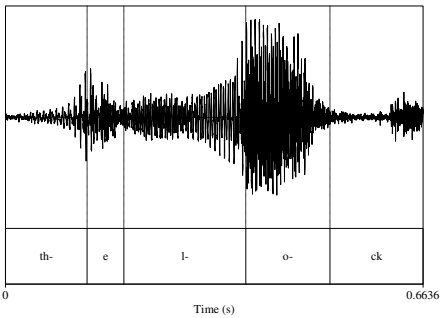
Example #2 (Informant #1)

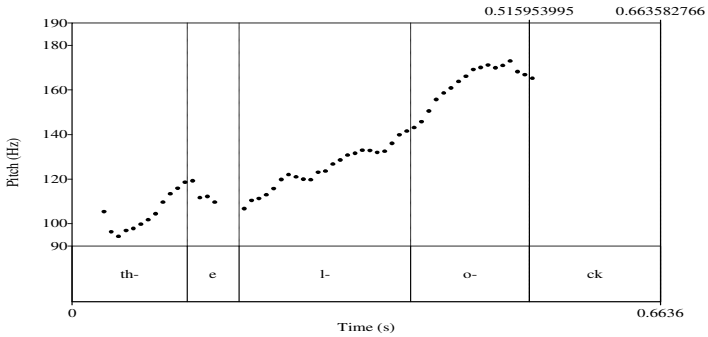


The pitch accent of Example #2 has been marked as H*+0, rather than H*+H; however, the sag effect during the pitch movement is clearly visible from the graph. The passage from "they" to "turn" occurs through an up-step, a feature which is so common that it could be even considered

a feature of GE: undoubtedly, the presence of an unvoiced consonant /t/ creates a break in the F0; when the consonant is changed into the vowel, the contour point is marked with limited delay. Analysing the graph, one can notice the evident curve which characterizes the pitch accent, even though the starting and final point have the same pitch height.

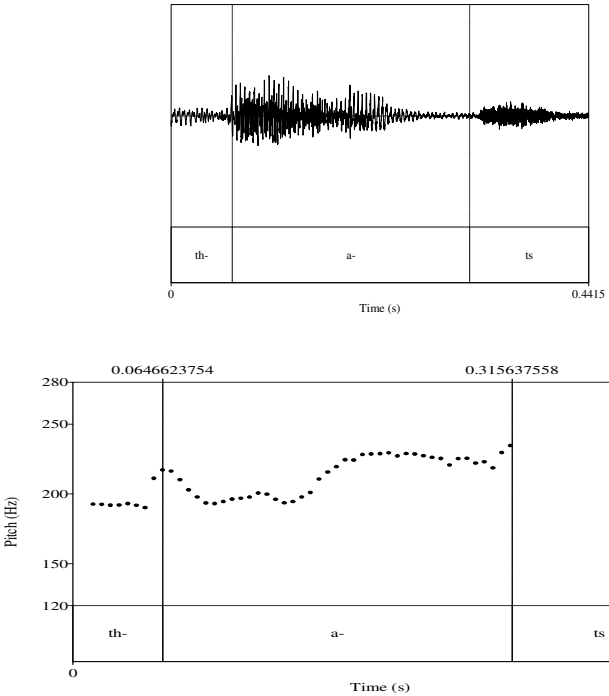
Example #3 (Informant #4)





Example #3 is extremely useful because it focuses on the role of liquids and semivowels. What is important to consider is how the /l/ has been realized in this H*+H pitch accent: as one can easily notice, the consonant has been kept for more than 200 ms; such is the intensity of the /l/ that PRAAT has even recorded some pitch movements. However, in spite of this feature, the H* is still perfectly aligned with no delay at all: when the consonant is changed into the vowel, the contour point is immediately marked. It is also interesting that the rise following the H* is rather swift.

Example #4 (Informant #2)



The focus of the utterance in Example #4 is "*that's*", which carries an H*+H pitch accent. Even in this IP, the movement towards the MPSE is rather rapid, although the dental fricative /ð/ is prolonged for almost 150 ms. When the consonant is changed into the vowel, the first contour point is immediately marked: it can be said that there is no delay in the H*, despite the rise that has shifted the pitch to a higher position. When the movement towards the stressed syllable is complete, the following rise is rather delayed, and it is not started before 90 ms. into the syllable.

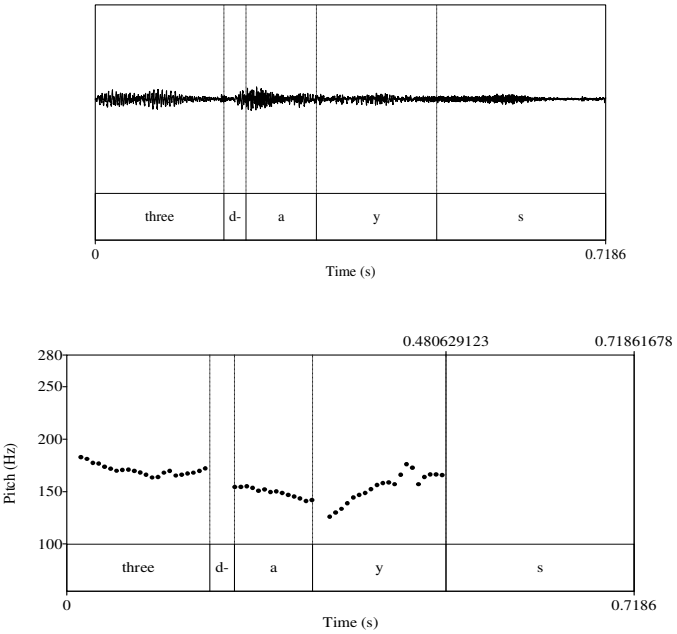
The graph reveals that the dental fricative has produced some sound which has been recorded by PRAAT. The question is whether the /ð/ can be compared to liquids and semivowels, which are often responsible for the rising movements towards the stressed syllable. However, the answer must be negative: while the pitch lines created by semivowels and liquids seem to have a logical continuation when /w/, /j/, /r/ and /l/ are changed

into vowels, those registered during dental fricatives do not appear to be completely consistent with the actual pitch line.

3.2. Rise on the second contour point (L*+H)

After presenting the H*+H and the H*+0 pitch accents, let us shift our attention to the L*+H. Despite the different first contour point (falling rather than rising), this pitch accent presents a final rise, too; as a result one could expect some similarities between H*+H and L*+H.

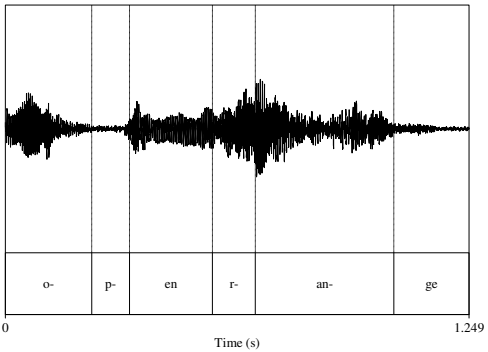
Example #5 (Informant #3)

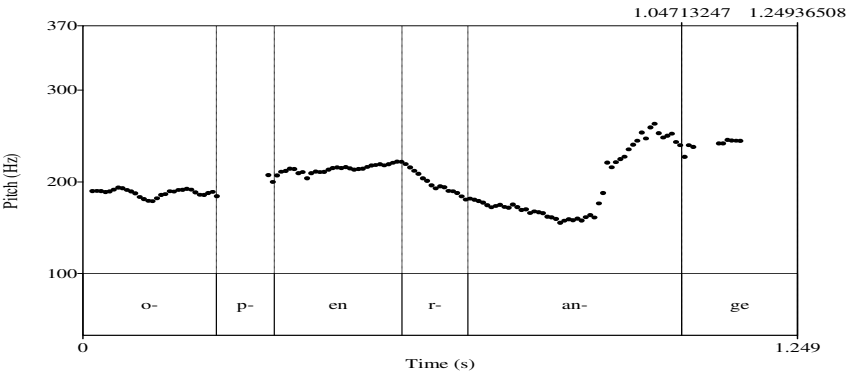


As previously explained, one of the main differences between the upward first contour point in CE and in GE is the alignment of the intensity peak in H*+H pitch accents: while in GE it is marked at the beginning of the syllable, in CE it is postponed and delayed.

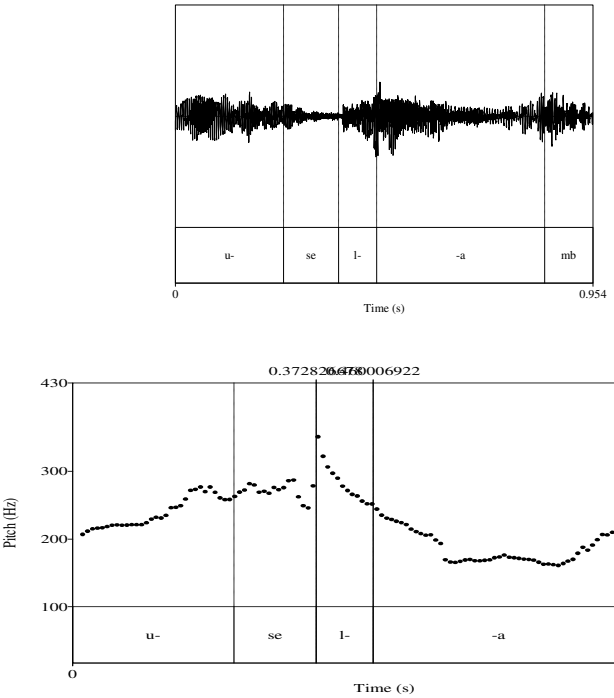
Example #5 seems to confirm that this pattern is also found in L*+H. The graph is very clear and shows that the intensity peak is reached the moment the /d/ is changed into the /e/, and so there is no delay in the alignment of the L*. When the first contour point is reached, the pitch continues to fall for another 108 ms. before it changes direction and moves upward.

Example #6 (Informant #1)





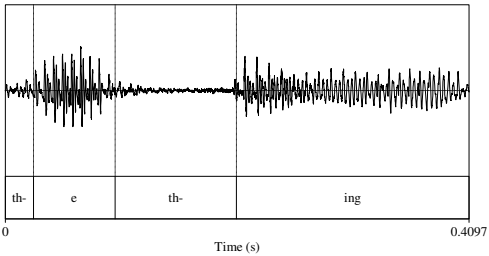
Example #7 (Informant #1)

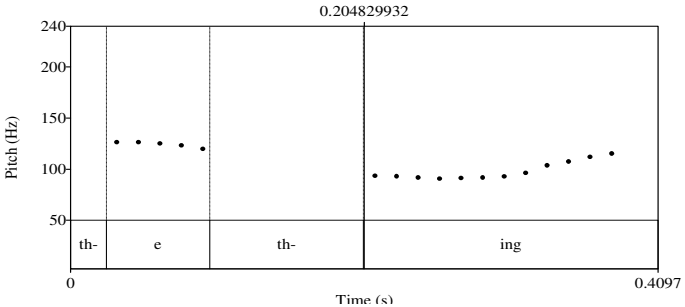


Example #6 (“range”) and Example #7 (“lamb”) are excerpts from the same interview and they present similar characteristics, and thus they will be presented together. In both pitch accents, the downward movement is carried by the liquids, /r/ and /l/, with the L* being marked when the two consonants change into the vowels. In both pitch accents, the new contour point does not lead to a change of direction, since the pitch continues to fall; the rise, instead, is introduced by a long static tone (around 90-100 ms.).

Example #8, which focuses on “thing”, presents a very similar structure: after the long initial fricative /θ/, the pitch completes its downward movement, with the L* being marked the moment the consonant changes into the vowel; the upward movement is preceded by a long static tone (100 ms.).

Example #8 (Informant #5)





3.3. The 0+H pattern in GE

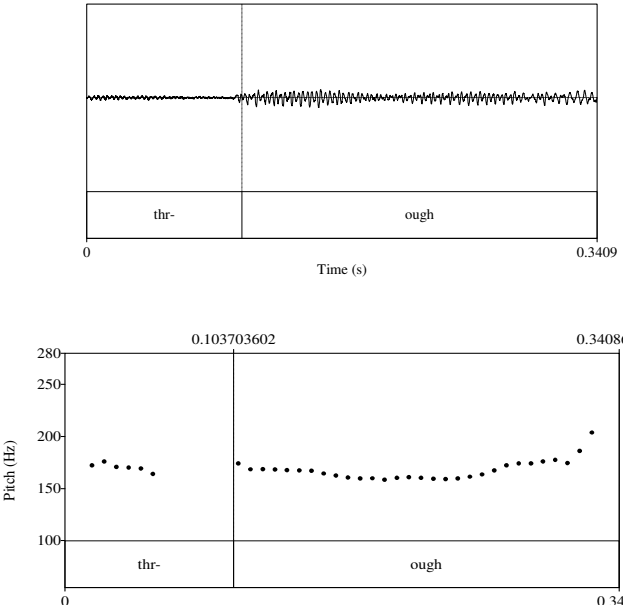
One of the most characteristic patterns of GE (as opposed to CE) is the 0+H, which consists of an extremely long level tone preceding the final rise; usually, it is introduced by a level first contour point (so, it is marked 0*+0+H), but there have also been cases of H*+0+H and L*+0+H. Since this pattern has been systematically found in almost all the investigated villages of Gwynedd, it is very useful to describe its main characteristics.

The natural position of the 0+H pattern is found before a boundary, which can be either major or minor; terminal tones are usually drawn out and lengthened, so one long tone can be expected. A second interesting issue is related to the MPSE status: the 0+H is linked to words and elements which the speaker wishes to highlight.

Despite having a final rising tone, this pitch accent gives the utterance a sense of conclusion, so it is extremely common in statements and declaratives; never has the 0+H pattern been found in questions nor has it

been used to express doubt or uncertainty. Since the 0+H pattern is found in monosyllabic words which contain either a vowel or a diphthong, the following analysis will be divided into two parts: the first will focus on monophthongs and the second on diphthongs.

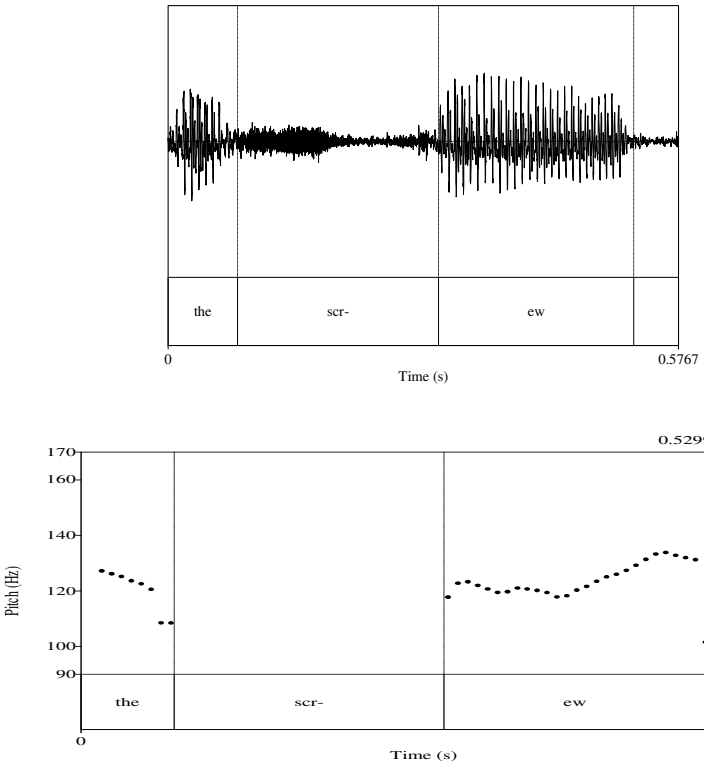
Example #9 (Informant #3)



Example #9 presents a 0*+0+H pitch accent on “through”. The first element which strikes the attention of the researcher is the long static tone which is developed during the pronunciation of the /u:/: after a lengthened /θr/, the pitch is kept at the same level for 210 ms., before it is allowed to rise; although the shape of the pitch line is not completely straight, movements are minimal. After the rise has begun, the pitch peak is reached after another 90 ms.

An interesting feature which characterizes the 0+H is the overall length of the word: the /u:/ lasts more than 300 ms., an amount which increases to 447 ms. if the initial consonant is included.

Example #10 (Informant #5)

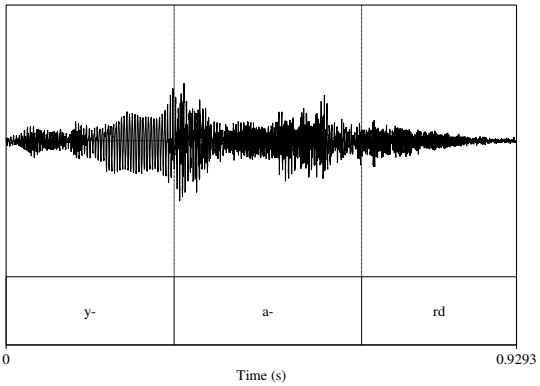


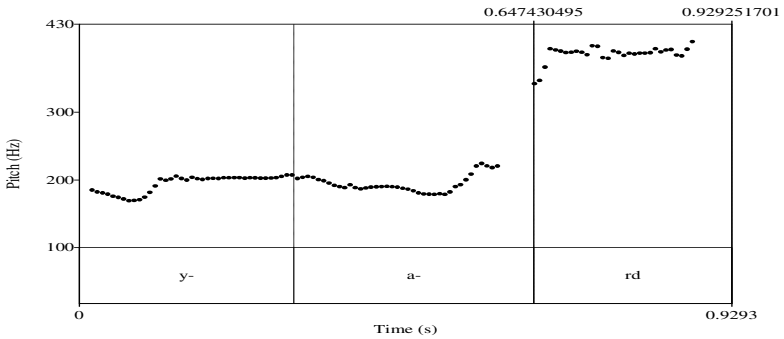
The pitch accent of Example #10 (“screw”) is rather similar to the one in Example #9, showing a long initial cluster /skr/ and the vowel /u:/. The length of the vowel is shorter than that in Example #9, but it still lasts 240 ms. (110 ms. for the static tone and 130 for the rise).

Example #11 presents an utterance with a long static tone: it begins during the semivowel, while the first contour point is marked when the /j/ changes into the /a:/. The pitch line during the full vowel is not even, but has a short decline just before the rise, as if it were a preparation for the upward movement.

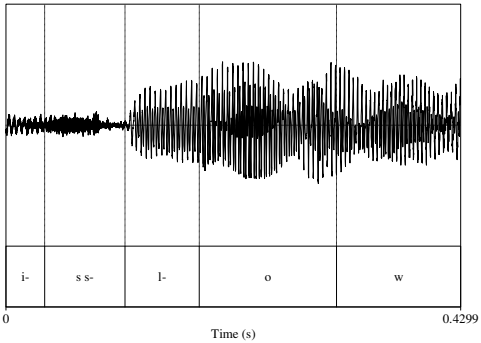
After focusing on monophthongs, let us now shift our attention to diphthongs. Examples #12 and Examples #13 present the usual features, so it will be not necessary to analyse them thoroughly. The graphs offer enough information about their pitch movements and intensity:

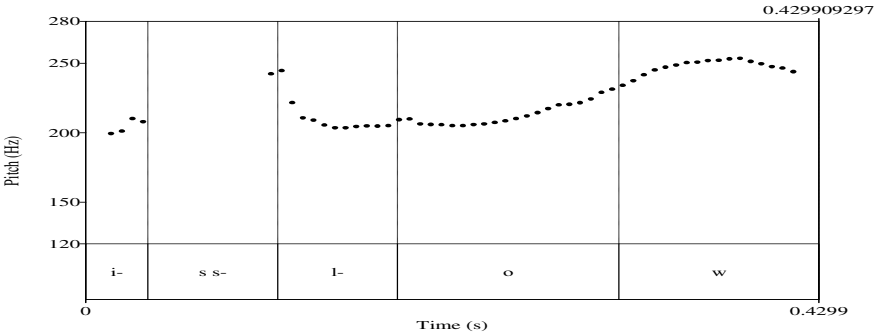
Example #11 (Informant #2)



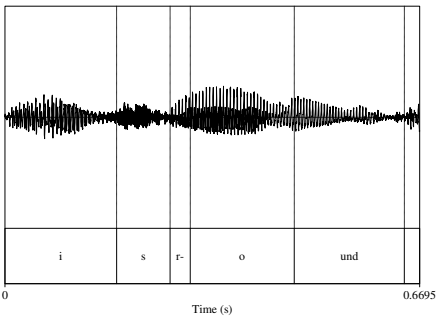


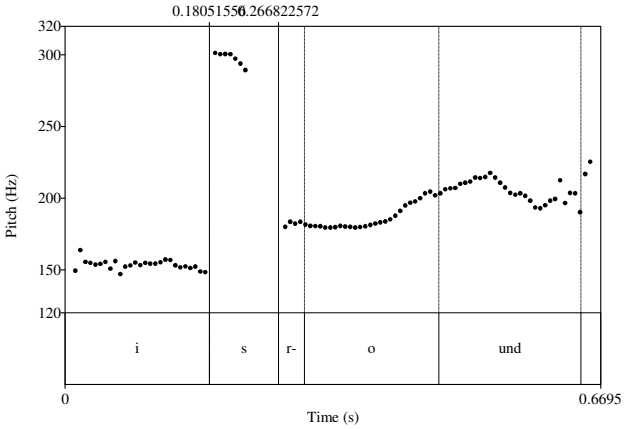
Example #12 (Informant #2)





Example #13 (Informant #2)





Both pitch accents ("slow", "round") present a long static tone (lasting no less than 110 ms.), followed by a gradual rise, which reaches its peak after about 100 ms. One short explanation should be given about the high pitch line in Example #13: the sibilant [s] has produced some sound which has been detected by PRAAT; however, it should not be considered in the pitch analysis.

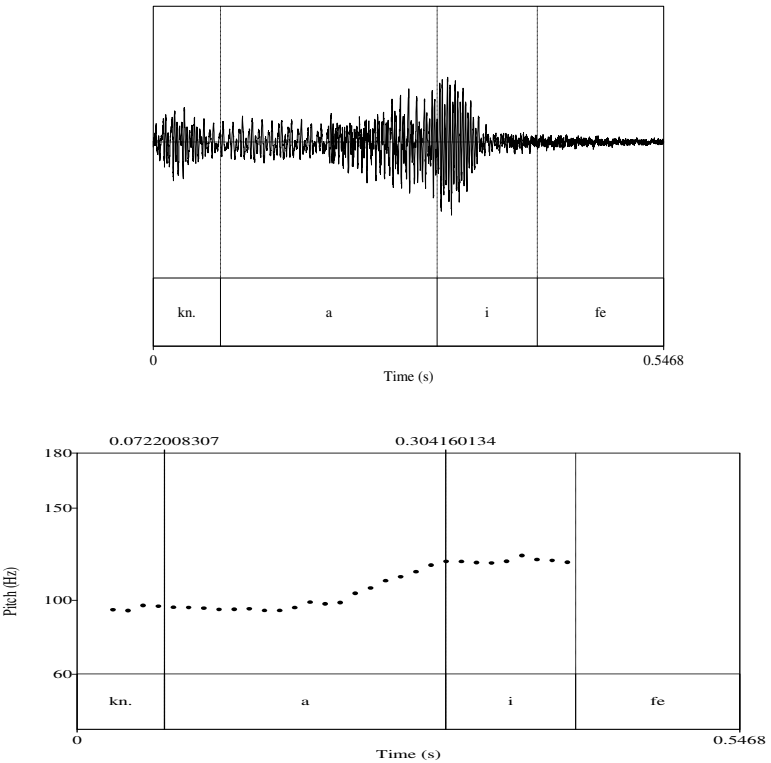
At the beginning of the chapter, two main features of 0+H were explained, namely the position of the pitch accent and its semantic role within the IP; after analysing several pitch accents, some other general and common characteristic can be added. Rises are usually gradual, although sometimes they can be introduced by a short fall, which should not be confused with sag. Upward movements are also very limited and seldom reach high values. There could be some debate about "yard" (Example #11), for example, since the pitch line in the graph seems to have reached a very high level at the end of the word; however, this should not be further considered, since it has been recorded during the pronunciation of two consonants /rd/.

The main issue is related to the intensity peak and the positions of the contour points, since there is not a complete uniformity. Usually, monophthongs have their intensity peak at the beginning of the syllable: sometimes this occurs when the consonant or the semivowel changes into the vowel, like in "screw of Example #10 and "yard" of Example #11.

Late alignment of the 0* or H* is also possible, but this is limited to no more than 40 ms.: this feature has been found in “through” in Example #9.

More complex is the situation of diphthongs: the first contour point, 0* or H*, can be aligned with a significant delay, which can even reach 100 ms., like in “slow” of Example #12. In addition, certain diphthongs present a preference for the second vowel, which is more marked and has greater intensity than the first: this feature occurs in Example #12 (“slow”), but also in Example #14 (“knife”)

Example #14 (Informant #6)

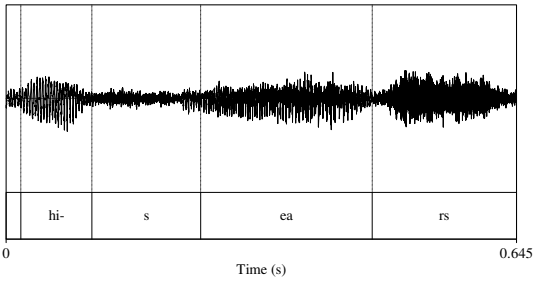


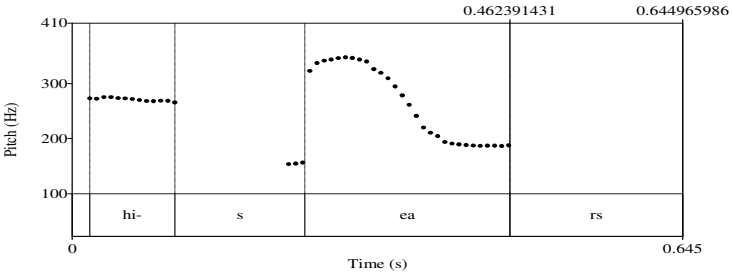
As a final note, it should be stated that the 0+H pattern has seldom been found in disyllabic words and it is never interrupted by a consonant.

4. Second contour point falling

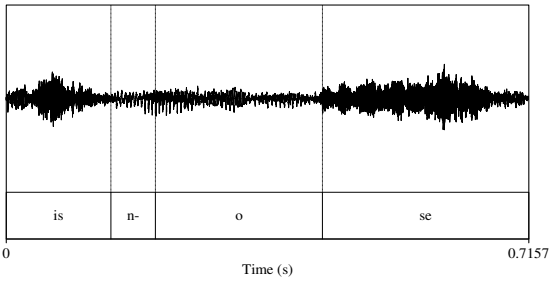
Let us now shift our attention to the H*+L. Example #16 and Example #17 are excerpts from the same interview and they share some similarities, so they will be analysed together.

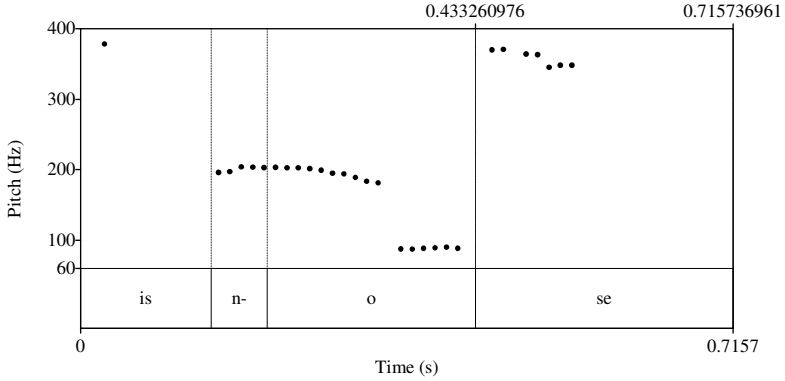
Example #16 (Informant #7)





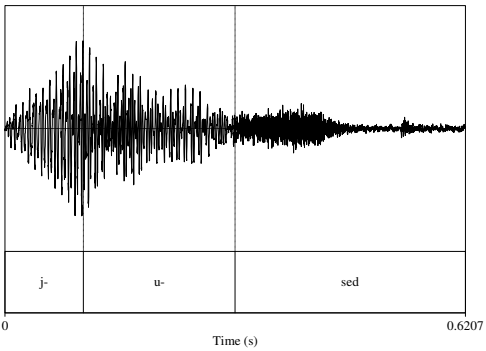
Example #17 (Informant #7)

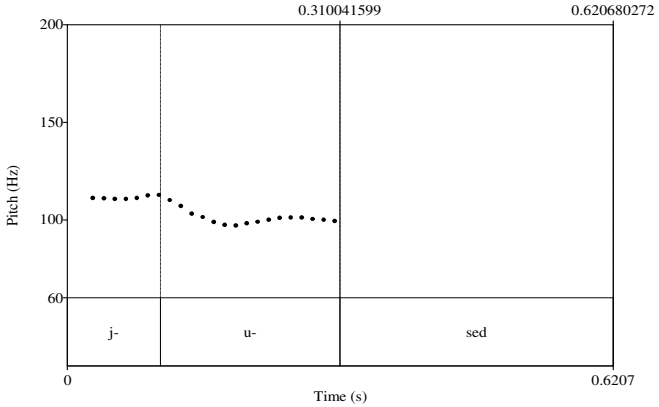




In both Example #16 and Example #17 the MPSE consists of a long vowel, which is kept for 219 ms. (“ears”) and 234 ms. (“nose”). Lengthening always causes a delay of the H*, and this feature appears evident in the two pitch accents. Before the marking of the first contour point, there is an evident static tone, which is the reason for the late alignment. However, even in H*+L pitch accents, the static tone always occurs after the rise and never before.

Example #18 (Informant #6)

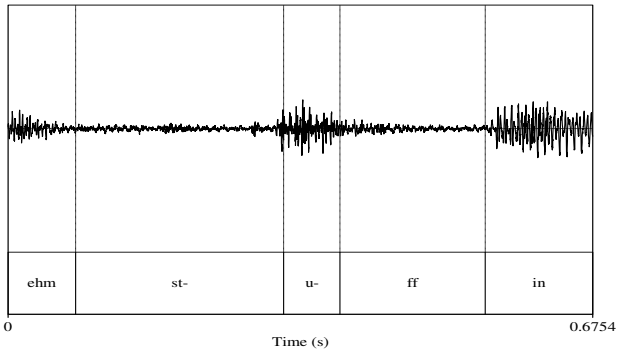


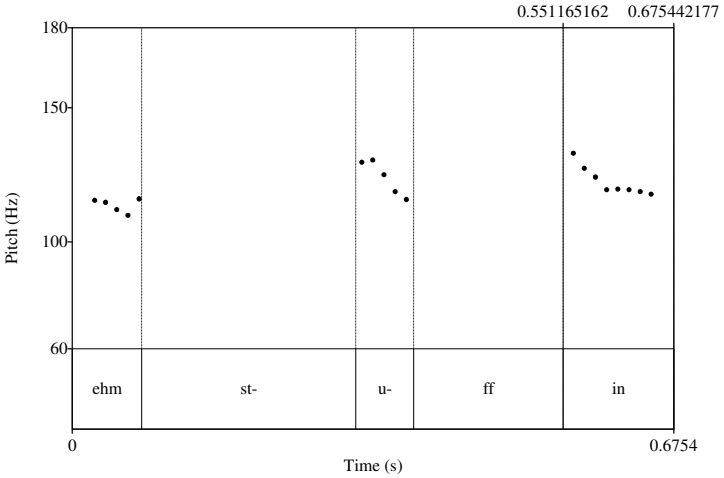


Example #18 also presents an MPSE with a long vowel. As has previously occurred, there is a delay in the H*, which is marked about 100 ms. after the beginning of "used". The pitch accent of Example #18 is characterised by two common features:

- The movement towards the stressed syllable is carried by a semivowel /j/
- The static tone starts and finishes at the same level of the H*, when the rise has already been completed

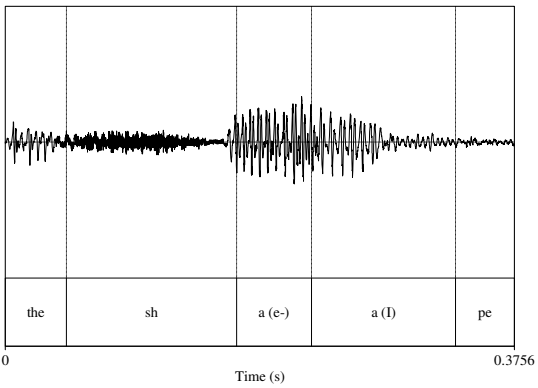
Example #19 (Informant #8)

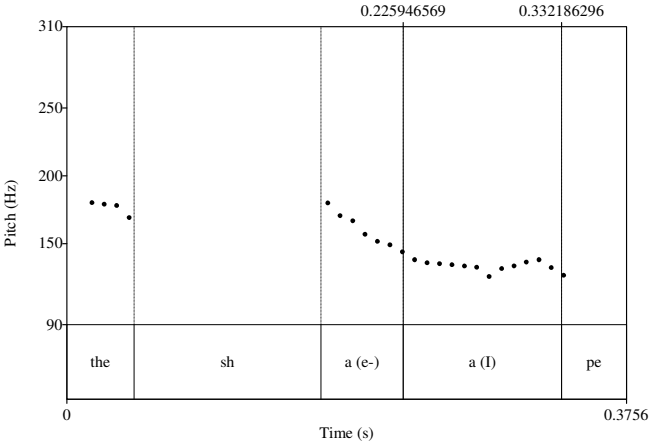




Even in "*stuff*" (H*+L), which carries a very short vowel, the alignment of the H* occurs without any delay.

Example #20 (Informant #2)





Example #20 will focus on the "shape", which has been described as an H*+L pitch accent: the diphthong /eɪ/ is pronounced after the long fricative /ʃ/. Different from the previous example, the two vowels of the diphthong have been separated. Although the greater intensity occurs on the /ɪ/, the /e/ is aligned immediately after the consonant, with very limited delay.

These two features are extremely important: similar research on other varieties of Welsh English, such as Ceredigion English (CE), has revealed that sagging effects or delayed rise are rather common. However, there is one important difference to remark on: while in GE the first contour point is aligned at the beginning of the syllable, in CE it can present a much bigger delay (even 80 ms.).

5. Pitch peak

The final part of the analysis on pitch accents will focus briefly on the pitch peak: here the goal is to determine the position of the highest pitch within an utterance; however, since substantial information on this feature has already been given above, this section will merely serve as a short review.

The first step is to present Walters' findings (1999, Chapter 5, Section 4.4.) on RVE: in H*+H, 0*+H and L*+H, the H-peak is reached rather late in single-syllable rises (79% into a long vowel, 92% into a short vowel);

when more syllables are involved, the peak is usually reached during the second syllable, but sometimes also on final nasals or liquids /l/.

The analysis of GE (and CE as well) has presented very similar data. In 0*/ H*/ L*+H, the pitch peak is reached towards the end of the word, with a percentage which can vary between 80% and 90% into the syllable[1]. The 0*+0+H is a very long pitch accent which also presents a final rise, thus the peak is reached even more than 90% into the syllable. In monophthongs, the pitch peak is reached at the very end of the syllable, sometimes even on the consonant after the vowel; in diphthongs, the peak can be reached earlier, with the tone being kept for some ms. at the highest positions.

Pitch peak and intensity peak do not always occur at the same moment: on some occasions, the final H is heavily stressed and more intense than the previous contour point, which leads to an almost perfect union between pitch peak and intensity peak. In certain other cases, the H is softer, the pitch peak being aligned much later than the intensity peak.

The comparison between the present results on CE and GE and those by Walters on RVE (1999) might suggest that the position of the pitch peak could be one general distinguishing feature of Welsh English in general; the confirmation of this hypothesis, however, could only be offered by future research on other varieties of WE.

References

Boersma, P. & D. Weenink, D. 2008. *PRAAT: doing phonetics by computer* (Version 5.0.32) [Computer program]. Retrieved 12th January 2010 from http://www.praat.org/.

Cruttenden, A. 1995. Rises in English. In *Studies in General and English Phonetics*, edited by J. Windsor Lewis, 155–173. London and New York: Routledge.

Gimson, A. C. 2008. *The Pronunciation of English*. London: Hodder Education.

Ladd, D. R. 2008. *Intonational Phonology*. Second edition. Cambridge: CUP.

[1] The percentage is calculated from the start of the initial vowel of the first syllable, thus the possible preceding consonant is not considered.

Tench, P. 1990. The Pronunciation of English in Abercrave. In *English in Wales. Diversity, Conflict and Change,* edited by N. Coupland, 130–141. Clevedon: Multilingual Matters.

Walters, J. R. 1999. *A study of the segmental and supra-segmental phonology of Rhondda Valleys English*. Unpublished PhD Thesis, University of Glamorgan, Pontypridd, Wales. [Condensed version published on-line by the University of Glamorgan. 2006. Retrieved 10th January 2011 from http://resnt1.isd.glam.ac.uk/rhondda_valleys_english/.

AN ACOUSTIC ANALYSIS OF THE PUNCH LINES IN ENGLISH JOKES

KEN-ICHI KADOOKA

Outline

The punch line Paratone is a subtype of the low Paratone observed at the end of jokes. It is a combination of the phonetic features of lower pitch, slower tempo in the punch line and a pause before it. By these features, the end of stories is signaled. In addition to these within-the-clause characteristics, the punch line Paratone includes the transition of the baseline pitches from the beginning towards the end of the story; high in the beginning and then the gradual lowering until the punch line. With the acoustic analysis done in this paper, these phonetic phenomena have been exemplified to a certain extent.

1. Introduction

This is an acoustic analysis of the Punch Line Paratone observed as a discourse intonation pattern. The punch line paratone is suggested in Kadooka (2009, 2011a, b) as a subtype of the Low Paratone. Tench (1996) and Wennerstrom (2001) define the Low Paratone as the temporary deviation from the main topic. Phonetically, the punch line paratone shares the following six observations with the Low Paratone (Tench 1996: 24, emphasis by the present author):

(1) 1.The high pitch on the onset syllable of the initial intonation unit.
2. The relatively high 'baseline' of that initial unit; this means that the low pitches are relatively high, compared to the low pitches in the final unit of the paragraph.
3. There is a gradual lowering of that baseline until the final unit is reached.

4. The depth of fall in the final unit is the lowest in the whole paragraph.
5. There is *usually* a slowing down process in the final unit.
6. There is a longer pause than is normally allowed between intonation units.

Let us label these with shorter headings as follows; each one corresponding with those above:

(2) 1. high in the beginning
2. high baseline
3. gradual lowering
4. lowest fall before the final
5. slower tempo in the final
6. longer pause before the final

In the sections below, we will use these headings. Of these six, the *baseline* of the second statement is slightly ambiguous. That is to say, the *baseline* of an intonation unit seems to indicate a level line of the pitch contour. In any case, the baseline of an intonation unit must be clearly defined. Given the restricted length of this paper, detailed discussion will be given on a later occasion.

The emphasized two adverbs *relatively* and *usually* are considered to be ambiguators or obscurers in the sense that they do not define something absolutely. In other words, the conditions listed in (1) allow some exceptions. When we simplify, somewhat extremely, the six conditions are modified as follows:

(3) 1. The high pitch on the onset syllable of the initial intonation unit. (the same as in (1))
2. The average pitch in each line is the highest in the initial line.
3. There is a gradual lowering of the average pitch toward the final unit.
4. The average pitch is the lowest in the final unit.
5. The tempo is the slowest in the final line.
6. The pause is the longest before the final line.

The reason for the modification into (3) is that the ambiguousness in (1) may be diminished with these simplified and extreme measures. These extreme figures will be easier to identify than those in (1).

Thus, the purpose of this paper is to acoustically exemplify whether the relative (1) or the absolute (3) is more suitable for authentic recordings of English jokes. To achieve this purpose, the acoustic analysis software *Praat* will be adopted.

As a general discourse structure, jokes are defined as follows:

(4) 1. Basically, jokes are told as a dialogue between two people.
2. The dialogue starts with a high tone of voice, which signals the beginning of the story.
3. The baseline of the tone of voice is gradually lowered toward the punch line.
4. Before the punch line, there must be a short pause to signal the end of the story.
5. At the punch line, the tone is the lowest in the whole story.

Of these five illustrations, 2 through 5 are concerned with the intonation pattern – especially the pitch variation.

2. English jokes

We will acoustically analyze four English jokes in this section. These jokes are adopted from Kadooka (2003), which is an English textbook for university and college students, hence the tempo is slower than the natural one, and the pauses between the intonation units are longer than usual. Recordings of these jokes were made by a British speaker.

There are four episodes to be analyzed: *I don't know, Discipline, Piggy Bank,* and *One Way Street.* The first two are dialogues between the characters, and the other two are not. The latter two are marked in the sense that jokes are basically a conversation between two people (see (4) above).

Below is a table showing the details of the acoustic figures of each story. The first column "ave." is the average frequency of the intonation group; the second one "min." stands for the minimum or the lowest frequency during the intonation group; "max." is the maximum or the highest frequency, and the "range" is the gap between the maximum and the minimum frequencies. All frequencies are indicated in Hertz. "Duration" is indicated in seconds, and "pauses" mean the interval before the line indicated by the second. The lowest figures of the average pitch, the highest through all the intonation units, the narrowest ranges and the longest pauses are emphasized with bold font in the tables. In the second

tables, the numbers of words, syllables (♦), the stressed syllables (feet) will be given, together with the duration time divided by each number: i.e. tempo. The largest figures of the duration – i.e. the slowest tempo – will be shown in the bold type.

The first story is the shortest of the four. It is a dialogue between a teacher and a student:

(5) Episode 1: *I don't know.* (7.15 seconds)
1 Teacher: Tell me three words most commonly used by students.
2 Student: I don't know.
3 Teacher: Correct.

Table 1. Acoustic figures and tempo for (5)

line	ave.	min.	max.	Range	duration	pause
1	140.74	79.73	204.87	125.14	3.78	
2	214.24	100.79	**265.93**	165.14	1.09	0.79
3	**115.10**	104.54	125.21	**20.67**	0.53	**0.81**

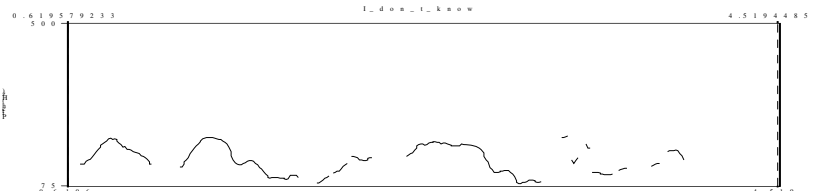
line	words	Tempo	s	Tempo	feet	tempo
1	9	0.42	12	0.32	5	**0.76**
2	3	0.36	3	**0.36**	2	0.55
3	1	**0.53**	2	0.27	1	0.53

This first episode is the shortest of the four jokes, consisting of only three turns. The first and the third turns are taken by the teacher, while the second one is taken by the student. The reader intentionally adopts high pitch to imitate the student's younger voice. Most characteristic of this episode is that the punch line consists of only one word: "Correct." In this punch line, the average pitch is the lowest, the range is the narrowest, and the pause before this is longer than the first one between turns 1 and 2. Another point to be mentioned is the tempo of the punch line. When we count the number of the words in each line, the tempo is the slowest in the punch line; one word for 0.53 second against three words for 1.09 second in line 2. But when we count by the number of syllables, two for the punch line against three for line 2. The third possibility is to count the number of stressed syllables, two for line 2 and one for line 3. Still the result with 0.55 second for each stressed syllable in line 2, is slower than 0.53 second in line 3.

The highest pitch through the episode appears in line 2. This is because the speaker imitates the student's young voice, hence it is higher than the counterpart in line 1. This seems to be not in line with the generalization in (1). To be more precise, let us look at the intonation contour:

Figure 1. Intonation contour of (5)

(6)



tell me three words most **com**monly used by students

The most prominent word is *commonly*, and the pitch peaks scattering in *tell, three, commonly* and *used*. When we loosely interpret the generalization in (1), one of the peaks comes in the beginning, hence it is certain that there is a 'high pitch on the onset syllable of the initial intonation unit' ((1) above).

The second episode is a dialogue between a father and his friend.

(7) Episode 2. Discipline (23.24 seconds)
1. The father was explaining to his friend the difficulty of trying to discipline his son. ↘
2."When I was his age, my father sent me to my room for punishment. ↘
3. But my son has his own TV, CD player and telephone." ↘
4."So what do you do?" →
5. asked the friend. ↘
6. ⇑ "I send him to my room." ↘

Table 2. Acoustic figures for (7)

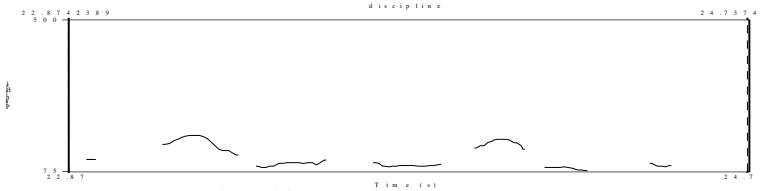
line	average	minimum	maximum	Range	duration	pause
1	128.49	82.96	182.20	99.24	5.24	
2	124.71	75.88	182.31	106.43	4.45	1.01
3	136.96	78.48	**188.29**	109.81	4.90	1.17
4	157.78	85.84	172.83	86.99	1.17	**1.33**
5	**86.33**	77.26	91.29	**14.03**	0.91	0.00
6	120.82	76.81	176.65	99.84	1.97	1.09

In this example, none of the lowest average pitch, the location of the highest pitch, the narrowest range, the longest pause obey the norms defined in (3) in section 1. The lowest average pitch is realized in line 5; the highest pitch in line 3; the narrowest range in line 5; the longest pause before line 4. The lowest average F0 and the narrowest range appear in line 5, because of the shortness of this line (three words) and the emptiness of the content. These three words are only to signal the speaker of the line, not to present any new information. If we exclude this line from the account of the lowest average F0 considering the latter factor, it is in the last line that the lowest average frequency is realized. The average pause between the lines is 0.92 seconds. The pause before the punch line is 1.09 seconds, longer than the average.

As for the tonic on *my* in the punch line (line 6), we must look at the F0 diagram:

(8)

Figure 2. Intonation contour of line 6 in (7)



The highest point (176.65 Hz) within this line appears in *send*, not in the tonic *my*. The prominence in this line falls on *my*, not on *send*, however, because the difference must be emphasized between the father's room and the son's room. This distinction by the prominence on *my* is a function of correction of Tone 5 (rise - fall) described in Halliday (1967, 1970, 1994) and Halliday and Greaves (2008).

The following is a table showing the tempo in each line of this story:

Table 3. Tempo in (7)

(9)

line	words	Tempo	σ	Tempo	feet	Tempo
1	15	0.35	24	0.22	6	0.87
2	14	0.32	17	0.26	5	0.89
3	11	**0.45**	16	0.31	4	**1.23**
4	5	0.23	5	0.23	1	1.17
5	3	0.30	3	0.30	2	0.46
6	6	0.33	6	**0.33**	2	0.99

When counted by the number of the syllables of each line, it is the slowest in the punch line (0.33 seconds per syllable in line 6). When counted by those of the words and the stressed syllables, however, it is the slowest in line 3 (0.45 seconds per word, and 1.23 seconds per foot). Either way, it is certain that the punch line is told with a slow tempo.

The next one is a dialogue between a mother and a daughter with a narration.

(10) Episode 3: *Piggy Bank* (30.11 seconds)

1. Mother decided that her daughter Judy, a nine-year old, was old enough to have her own bank account.
2. So she took Judy to the local bank.
3. Judy liked this idea very much.
4. "This is to be your account.
5. You must fill out the application yourself."
6. Judy did well until she came to the space marked, "Name of your former bank."
7. She thought for a while and wrote down,
8. "Piggy."

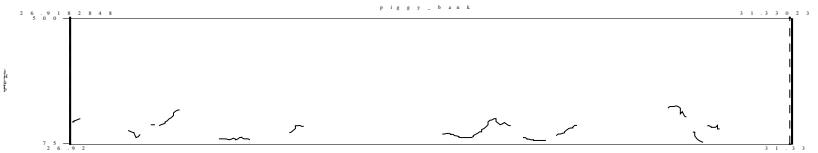
Table 4. Acoustic figures for (10)

line	average	minimum	maximum	Range	duration	pause
1	123.17	75.06	204.39	129.33	6.65	
2	136.29	94.40	**207.45**	113.05	2.21	0.70
3	117.19	80.96	192.38	111.42	2.37	**1.24**
4	126.67	81.20	204.52	123.32	1.81	0.84
5	125.09	89.33	177.20	**87.87**	2.54	0.57
6	125.45	86.92	199.23	112.31	5.54	1.00
7	115.65	85.15	188.92	103.77	2.90	0.87
8	151.91	82.25	202.41	120.16	0.47	0.40

Below is the pitch contour of lines 7 and 8:

Figure 3. Intonation contour of lines 7 and 8 of (10)

(11)



she **thought** for a while and wrote down **pig**gy

The average frequency in the punch line (151.91 Hz) shows the highest through the story, rather than the lowest. This may be because this line is 'told' by a nine-year old girl - actually she wrote down 'Piggy,' it was not uttered by her. The narrowest range does not appear in the punch line, nor the longest pause before it. The highest pitch appears in the second line, in addition. Thus, this story is an exception to the standard defined in (3) in section 1. This irregularity may come from the idiosyncrasy that lines 4 and 5 are the mother's turns but Judy does not say anything herself. "Name of your former bank." in line 6 is what is written on the application form, and the punch line "Piggy" is the answer that Judy wrote down on that application form. As an auditory impression of the recording, however, this punch line sounds as if it was uttered by Judy.

The analyses of tempo in this episode is as follows:

Table 5. Tempo in (10)

(12)

line	words	tempo	Σ	Tempo	feet	Tempo
1	18	0.37	25	0.27	8	0.83
2	8	0.28	10	0.22	2	1.11
3	6	0.40	8	0.30	3	0.79
4	6	0.30	7	0.26	2	0.91
5	7	0.36	11	0.23	3	0.85
6	15	0.37	18	0.31	6	0.92
7	8	0.36	8	**0.36**	3	**0.97**
8	1	**0.47**	2	0.24	1	0.47

The average pause between the lines is 0.80 seconds, and the pause before the punch line is 0.40 seconds. Thus, the pause before the punch line is shorter than the average. Though the tempo is the slowest in line 8 when counted by the number of the words, it is line 7 when counted by the numbers of syllables and feet. The reason for this disagreement should be ascribed to the idiosyncrasy of this episode, in that the punch line consists of only one word. Hence the duration per word is the longest in this line, but it is not when counted by the numbers of syllables and feet. As an auditory impression, however, the recording of the punch line sounds slow.

The last episode is the longest one of the four. The whole story is told by narration, including the punch line written on a piece of paper; it is NOT told by a character.

(13) Episode 4: *One Way Street* (36.11 seconds)

1 A Frenchman was visiting New York for the first time in his life.
2 He could not speak English at all.
3 One day he decided to go for a walk.
4 He was afraid of getting lost,
5 so he carefully looked at the street sign in front of his hotel
6 and wrote it down on a piece of paper.
7 He walked around for quite a long time.
8 Then he realized he was lost.
9 He saw a policeman
10 and showed the piece of paper.
11 It said
12 "ONE WAY STREET."

Table 6. Acoustic figures for (13)

line	average	minimum	maximum	Range	duration	pause
1	122.49	92.20	207.86	115.66	4.36	
2	122.74	87.72	156.00	68.28	2.28	0.90
3	140.79	85.00	193.08	108.08	2.48	0.81
4	118.53	78.34	156.19	77.85	1.99	0.90
5	132.07	87.82	204.83	117.01	4.15	0.45
6	124.73	88.34	185.10	96.76	2.24	0.37
7	130.71	75.40	**210.59**	135.19	2.61	**1.38**
8	133.05	85.60	196.89	111.29	2.08	1.34
9	138.40	86.80	186.85	100.05	1.22	0.53
10	118.86	91.85	154.57	**62.72**	1.71	0.41
11	128.25	101.89	176.49	74.60	0.69	1.34
12	117.67	82.18	199.48	117.30	1.38	0.49

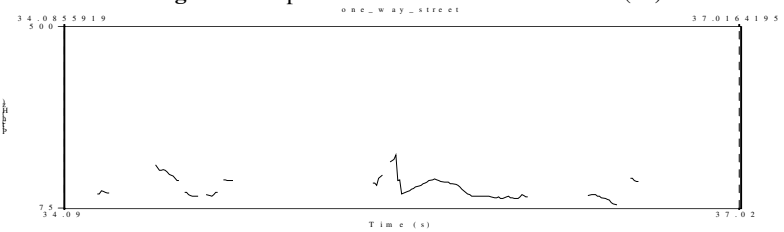
The lowest average frequency is found in the last line, which is in line with the ideal pattern of pitch transition. The other factors, however, do not coincide: the highest pitch is in line 7, the narrowest range in line 10, the longest pause before line 7.

This story is an exception to the discourse structure of jokes illustrated in (4) in section 1 in that the whole episode consists of narration, not speeches by the characters. Though there are two characters in this story, namely a Frenchman and a policeman, neither of them speaks. The punch line "ONE WAY STREET" is what is written on a piece of paper; but still there is a possibility that the police officer read that out.

Below is the pitch contour of lines 11 and 12:

(14)

Figure 4. the pitch contour of lines 11 and 12 in (13)



It said, **one** way street

As we see in this picture, line 11 '*It said*' is a fall-rise pattern, while line 12 'ONE WAY STREET' is a rise-fall. The tonic of these two lines is on 'ONE' in line 12.

The analyses of tempo is as follows:

(15)

Table 7. Tempo in (13)

line	words	tempo	σ	Tempo	Feet	tempo
1	13	0.34	16	0.27	5	0.87
2	7	0.33	8	0.29	3	0.76
3	9	0.28	11	0.23	3	0.83
4	6	0.33	8	0.25	2	1.00
5	13	0.32	16	0.26	3	**1.38**
6	9	0.25	10	0.22	2	1.12
7	8	0.33	9	0.29	2	1.31
8	6	0.35	7	0.30	2	1.04
9	4	0.31	6	0.20	2	0.61
10	6	0.29	7	0.24	2	0.86
11	2	0.35	2	0.35	1	0.69
12	3	**0.46**	3	**0.46**	1	**1.38**

This is the most straightforward result of the four stories analyzed here, all of the three figures coincide in the punch line: the numbers of words, syllables and stressed syllables. The average pause between the lines is 0.81 seconds, and the pause before the punch line is 0.49 seconds. Thus, the pause before the punch line is shorter than the average. This is because lines 11 and 12 are to be regarded as one utterance *It said "ONE WAY STREET"* from a syntactic perspective, though they consist of two

intonation units. In fact, the pause before line 11 is 1.34 seconds and the second longest with the small difference of 0.04 seconds to 1.38 seconds before line 7.

Thus, the line-by-line analyses have been carried out in this section. In the next section, the more detailed analyses (across clause boundaries) will be presented.

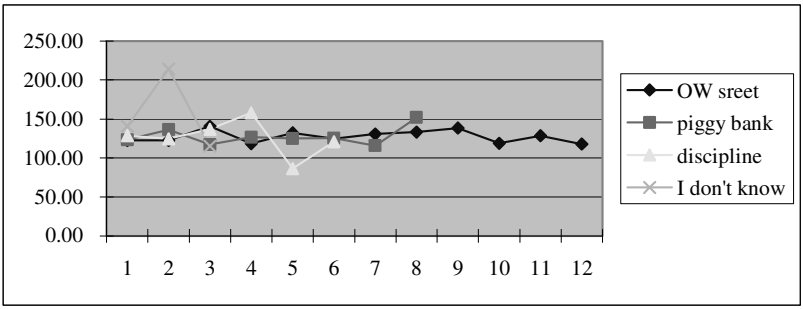
3. Gradual lowering

In this section, we will focus on the gradual lowering of the baseline, the third feature pointed out in (1) and (2) in section 1.

The transitions of the average pitch in each episode is as follows:

(16)

Figure 5. Transitions of the average pitch in each episode



As an observation among the stories, it can be pointed out that the more intonation units a given story contains, the more stable the transition within the story becomes. The most stable pattern is shown with the longest one (*One Way Street)*, while the least stable one is the shortest *I don't know*. The other two come in between these two extremes.

When we extend the transition of the pitch in each story, 'the gradual lowering of the baseline' in (1), statement no. 3 above can be paraphrased as follows:

(17) Each story begins with the highest tone of voice of the speaker, and terminates with the lowest tone; the tone gradually lowers during the story.

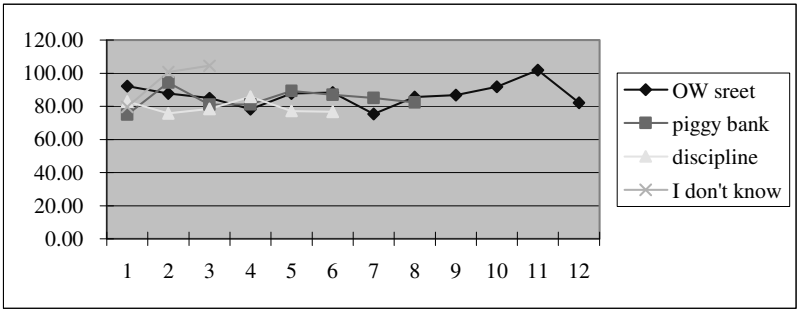
To simplify, it would be reasonable in a sense that the second highest pitch is found in the second line, the third in the third line, etc. The figure in *One Way Street* goes in line with this pattern the most; this story is told as a narration, not by the characters. All the lines are narrated by one speaker, and the transition is 'monotonous.'

It may be significant here that the punch line is what has been written on a piece of paper, not told by the Frenchman who cannot speak English at all. Though the situation is similar in that the punch lines are narrated and NOT told by the character themselves, the tone of voice is quite different in *Piggy Bank* (10) and in *One Way Street* (13); it is a nine-year old girl who speaks out before the punch line in *Piggy Bank*, and it is an adult male in *One Way Street*.

The reality in other stories is, however, far from this simplification. If the speaker should pretend to be a younger character such as a nine-year old girl in *Piggy Bank* or a student in *I don't know*, the voice must be higher than the ordinary tone. Hence, it is in line 2 in *I don't know* and line 8 or the punch line in *Piggy Bank* that the highest pitch appears. To summarize, the appearance of the highest pitch depends on the discourse structure in the sense that characters such as children and females requiring a higher voice do not always appear in the first line.

When we look at the baseline, or the low pitches instead of the average pitches, what will the result be? Below is the transition of the low pitches of the four episodes:
(18)

Figure 6. Transition of the low pitches of the four episodes



The most probable candidate to fit in the definition of the gradual lowering may be *Piggy Bank*, when we look at (18). Though the highest

pitch is in the second line, we can find a declining tendency toward the eighth line, especially from the fifth line. The changes in *Discipline* are close to this lowering pattern. The other two (*I don't know* and *ONE WAY STREET*) are, however, NOT close to the generalization of the gradual lowering in (1) and (2). The change in *I don't know* is the opposite of lowering; it is a gradual rising. As pointed out in the previous section, the second line is told as if it were a student; hence the lowest pitch in the second line is higher than that in the first line. As for the changes in *ONE WAY STREET*, it can be pointed out that there is a lowering in the first half until line 7, but in the second half the tone is rising from line 7 to line 11.

To summarize, it would be difficult to confirm to what extent the generalization of the gradual lowering is applicable, from our examination of the pitch patterns of the four English jokes. One of the main reasons for such exceptions may be idiosyncratic to each story. To take an example in *Piggy Bank*, the punch line 'Piggy' is told by a nine-year old girl. From the phonetic-semantic viewpoint, it seems necessary to define what kind of meaning and function the average and the lowest pitches have.

4. Conclusion

So far, we have looked at the acoustic analyses of some English jokes, with the special focus on the pitch contour and Paratone. In this section, we will attempt to find out what is closer to the truth: the auditory impression of Tench (1), or the extreme acoustic version (3).

Below is a table summarizing the analyses in section 2. The numbers 1, 3, 4, 5, and 6 correspond to those in (1) and (2), and indicated in each cell are those of the episodes which satisfy the stated conditions:

Table 8. Summary of the analysis

(19)

	(1)	(3)
1. high in the beginning	1, 2, 3, 4	
3. gradual lowering	2, 3, 4	
4. lowest fall before the final	1, 2, 4	1, 4
5. slower tempo in the final	1, 2, 3, 4	1, 2, 3, 4
6. longer pause before the final	1, 2, 4	1, 2

At a first glance, the conclusion would be that the auditory impression in (1) seems superior to its extreme counterpart (3) in that more satisfactory

results are given with the former. When counted by the mere numbers of the episodes, 17 with (1) against 8 with (3).

Let us examine each item one by one. As for the high beginning, the results are contrastive in that all of the four episodes show agreement with the loose conditions in (1), but none of the four corresponds with (3). The results of the gradual lowering show a similar tendency in that three correspond with (1) but none with (3). With regard to the slower tempo in the final, the results do not differ between (1) and (3); all of the four episodes show agreement both in loose and extreme definitions. The other two, the lowest fall and a longer pause before the final, come between these in the sense that some of those are in agreement with (1) are not in (3).

To conclude, the description of Paratone in (1) is quite valid as the result of the acoustic verification of the English jokes. When they are interpreted extremely as in (3), the results are less valid, however.

References

Halliday, M. A. K. & W. S. Greaves. 2008. *Intonation in the Grammar of English*. London: Equinox.

Kadooka, K. 2009. Punch Line Paratone: A special use of discourse intonation. *The Ryukoku Journal of Humanities and Sciences* 31 (1): 203–216.

Kadooka, K. 2011a. An acoustic analysis of the Punch Line Paratone in English jokes. *The Ryukoku Journal of Humanities and Sciences* 33 (1): 1–13.

Kadooka, K. 2011b. A cross-linguistic study of Punch Line Paratone in Japanese and English. *Japanese Journal of Systemic Functional Linguistics* 6 (1): 1–15.

Tench, P. 1996. *The Intonation Systems of English*. London: Cassell.

Wennerstrom, A. 2001. *The Music of Everyday Speech*. Oxford: OUP.

OBSERVATIONS ON THE NUCLEUS IN ENGLISH AND SERBIAN

BRIAN MOTT

Outline

In this chapter I attempt to summarize some of the differences between English and Serbian utterances as regards the position of the nuclear stress.

Having first established that the versatility of the nucleus in Serbian is comparable to that of English by asking informants for Serbian equivalents of the sentence *Place the broken glass in the bin* spoken in a number of different versions created by changing the focussed element in each case, I then prepared a list of examples for recording, guided partly by Wells, *English Intonation* (CUP, 2006), to test which elements of the Serbian sentence can receive tonic stress and investigate the concomitant change in pragmatic value, and to determine the differences between English and Serbian in this respect.

Analysis of the recordings has enabled me to group the utterances into a number of types, such as WH- questions, interrogatives with an emphatic particle in Serbian, and those containing negative adverbs, intensifiers, topicalized objects, place names including words like *street*, *park*, etc., emphatic pronouns or possessives, comparatives, stressed prepositions, and so on.

Using a selection of examples taken from my corpus, the article will discuss the findings that emerged from comparing the English and Serbian utterances.

1. Introduction

My interest in sentence stress derives from having paid particular attention to the position of the nucleus in English while teaching English intonation to Spanish learners of English for many years (see Mott 2011, chapter 10). The work of Bolinger (1965, 1972, 1986, 1989), Cruttenden

(1990, 1997), Ladd (1980, 2008), Selkirk (1984, 1995, 2002) and Wells (2006) has proved to be especially useful in this area.

I have also given some thought to Spanish, a western Romance language considered to have the nucleus nearly always at or near the end of an utterance, and I have found cases where this is not so, which are discussed in Mott 2009 (see also Mott 1993). Likewise, I have examined the eastern Romance language, Romanian, of special interest for its heavy Slavonic and Hungarian overlay on the Romance stratum. Like these last two languages, Romanian tends to place the sentence stress on interrogative words and on the negative adverb *nu* 'no' in preference to lexical items in the utterance.

Therefore, it seemed that it was time to look at at least one Slavonic language in view of the versatility of the nucleus in this language family in addition to a very flexible constituent order, and the fact that the notion of the information structure of sentences, known as Functional Sentence Perspective, originated in eastern Europe in the tradition of the pre-war Prague School with scholars such as V. Mathesius, and was developed in the early 1960s by J. Firbas (see, for example, Firbas 2006), a disciple of J. Vacek. It is Vacek, in fact, who suggested not only the term Functional Sentence Perspective, but also Communicative Dynamism, both of which were adopted by Firbas (see Firbas 2006: xii & 104). For his part, Mathesius showed that the principal determinant of English word order is grammar, while in Czech the dominant role is played by the Functional Sentence Perspective linearity principle (Firbas 2006: 119).

2. Procedure

To check the mobility of the nucleus in Serbian, I used the sentence *Place the broken glass in the bin* and asked a native speaker to give possible Serbian versions of it as I shifted the tonic from one word to another (excluding the definite article) and explained the difference in meaning of each possibility for the native speaker of English. The test confirmed that the sentence stress can be on any element in Serbian, including the preposition, as in English, without there necessarily being any change of word order. Accented prepositions are not usually a possibility in some languages like Spanish and Romanian.

I then constructed 92 sentences based partly on Wells (2006) and asked my participant in the experiment how these might be performed naturally by a Serbian speaker, providing explanations of the possible meanings intimately related to the position of the nucleus where necessary. After

making an initial recording and finding that a number of the utterances were prosodically similar in English and Serbian, I subsequently produced another 65 sentences and made a further recording of the corresponding Serbian versions with a different speaker, which I hoped would sharpen the results of my inquiry.

The recordings were edited so as to eliminate versions of the utterances that the participants deemed to be unnatural or unsatisfactorily performed, and I later made an attempt to categorize the data. Even after editing, some sentences were still kept on the recordings in more than one version, these being sometimes alternative interpretations of the English versions, and other times renderings of the same utterance that were only slightly distinguishable from one another, but nevertheless worth examining. These multiple versions proved to be useful as a means of checking the reliability of what participants considered to be the "best" rendering of an utterance.

3. Results

In order to attempt a classification of the examples recorded in Serbian, I used some of the subdivisions proposed for Slovene in Šuštaršič (2005). In the cases in which more than one single version was recorded, I have tried to include additional comment where appropriate, especially where the differences appear to change the pragmatic value of the utterance. It must be remembered, however, that intonation and stress patterns are highly idiosyncratic, and more research with a greater number of speakers would no doubt increase the number of possible versions and would fine-tune my work.

In each of the sections below, the examples are given first in English, then in Serbian, and are preceded by remarks if it was thought that some clarification was required. The nucleus in each case is underlined as accurately as possible, though, in a few instances, the assignment of certain consonants to particular syllables may be controversial.

3.1. WH-questions

In broad focus WH- questions, Serbian places the nucleus on the WH-word. The situation is the same even when an emphatic particle like *pa* is introduced (section 3.1.1), unlike Slovene, in which this particle sends the tonic stress to the end of the utterance: *Kam pa je <u>šel</u>?* 'Where did he <u>go</u>?' (Šuštaršič 2005: 34).

The Serbian interrogative adverbs *zašto* 'why?' and *koliko* 'how much?', and the interrogative adjective *koji* 'which' do not necessarily take the sentence stress. On the other hand, *još* 'more' attracts nuclear stress, as do some prepositions, like *za* and *po*, both meaning 'for' (19–21; see also section 3.5). Pragmatics, of course, also plays a part in determining the position of the nucleus. Utterance 22 was pronounced with a high head + high fall in English in order to express greater surprise and, in this case, the Serbian equivalent revealed shift of the nucleus to the first syllable of *suđe* rather than assignment to *gde*. Similarly, if the nucleus were on *bilo* in example 3, instead of on *šta*, the utterance would imply more interest on the part of the speaker; in 4, if *zna* were accented instead of *ko*, it would render the utterance more contemplative.

1. Who's that? *Ko je to?*
2. What's that? *Šta je to?*
3. What's the matter? *Šta je bilo?*
4. Who knows? *Ko zna?*
5. Where have you been? *Gde si došao?*
6. How are you? *Kako si?* BUT How's it been? *Kako ide?*
7. Where is he from? *Odakle je?*
8. How long are we going to carry on like this? *Dokle ćemo ovako?*
9. What's that for? *Čemu ovo služi?*

3.1.1. Interrogative word with an emphatic particle in Serbian

10. Where the devil have you been? *Pa gde si došao?*
11. What are you doing? *Pa šta to radiš?* What the devil are you doing? *Šta bre radiš?*
12. Where are those books? *Pa gde su te knjige?*
13. When will they arrive? *Pa kad će već stići?*
14. Where can he be? *Gde li je?*

3.1.2. Interrogative word does not contain the nucleus in Serbian

15. Why didn't you wait? *Zašto nisi čekao?*
16. What time does the train leave? *U koliko sati polazi voz?*
17. How long have you been waiting? *Koliko čekaš?*
18. What size do you take? *Koja veličina vam je potrebna?*
19. What else do you need? *Šta ti je još potrebno? (Cf. Da li ti je potrebno još nešto?)*
20. How much is it? *Pošto je?*
21. What's it for? *Za šta je ovo?*
22. Where's the dishwasher. *Gde je mašina za suđe?*

3.2. The negative adverb *ne* in Serbian

The negative adverb *ne* in Serbian is normally the nucleus in broad-focus sentences, a fact that tends to mean that the sentence stress will be nearer the beginning of the intonation phrase than in English, except in the case of question tags as in 27. However, this rule can be overridden by narrow focus on another word (30–33).

23. Not necessarily. *Ne mora da znači / Ne baš.*
24. You can't do that. *Ne smeš to da radiš.*
25. You mustn't tell him. *Ne smeš da mu kažeš.*
26. There will be no money this year. *Ove godine neće biti novca. (… novca neće biti).*
27. Lovely day, isn't it? *Lep dan, zar ne?* (Question tag)
28. A: Could I borrow some sugar? B: I haven't got any. *A: Mogu li da pozajmim nešto šećera? B: Nemam šećera.*
29. A: Why are we going to Majorca again? B. We're not going to Majorca! *A: Zašto idemo ponovo u Majorku? B: Nećemo ići u Majorku.*

3.2.1. Cases in which Serbian *ne* is not the nucleus

30. He does come round, but not very often. *On navraća, ali ne često.* (Cf. Šuštaršič 2005: 37)
31. Five, not six. *Pet, ne šešt.*
32. I'm not so sure about that / I'm not quite sure about that. *Nisam baš siguran u to.*
33. It's not very good. / It is not very good. *Nije baš dobro.*

3.3. Other cases of fronted stress in Serbian and/or English

Note, in particular, that when Serbian puts the object in front of the verb in order to highlight it (3.3.2), it takes sentence stress, and the whole utterance is said in a single intonation phrase (IP), whereas English divides the utterance into at least two IPs (e.g. *Chocolate | I like*), except in the case of "cleft sentences" (e.g. *It's chocolate that I like / What I like is chocolate / Chocolate is what I like*). The structure is less usual and more marked in English than in Serbian.

So-called "event sentences" (3.3.3) require further study, because it is highly likely that whether the verb will contain the nucleus or not will depend on its semantic richness. Regarding Slovene, Šuštaršič (2005: 43–

44) limits himself to saying that it displays the same patterning as in English in this case or, in other words, the subject is selected for focus, although he admits that an object may also contain the nucleus if at the beginning of an IP. Regarding 44, note that the pattern in Slovene would always be *Dez pada*, never *Dez pada* (Šuštaršič, personal communication).

Names of streets, squares and stations, etc. (3.3.5) generally accent the name in Serbian, though note 51 (cf. Šuštaršič 2005: 36, *stari trg*); this only occurs in English in the case of combinations with "street", in which this word is not accented, the accent being on the name, as in Serbian (46).

Example 53 shows different patterning for Serbian and English since Serbian uses strong pronominal forms after prepositions and these may be accented. In many other cases of narrow focus (anaphora, emphatic personal pronoun, etc.), Serbian and English coincide.

Both English and Serbian deaccent anaphoric elements, but note that in the process Serbian cannot accent weak auxiliaries, only strong ones, so in example 55, the tonic still falls on the Serbian word for *sensible*, even if it is anaphoric, unless a strong auxiliary is used.

Both English and Serbian can bring personal pronouns into narrow focus (62–66), but notice that the Serbian construction in 63 corresponds to something like *Does that not please them?* in English, with emphasis on *that*. Šuštaršič (2005: 42) says that in cases where English places the nucleus on the pronoun, the nucleus goes on the verb in Slovene. Thus the Slovene equivalent of 62 would be: *Jaz pa sem jo VIDEL.*

The emphatic possessives in 3.3.8 require no special comment, but the examples in 3.3.9 do require further explanation. Why should English *day* and Serbian *dan* be the nucleus in 70 if it is obvious from the context that the concept *day* is already part of the background? Perhaps they are idiomatic expressions and have a fixed tonicity. Wells (2006: 112) suggests that *day* is accented because it hasn't been mentioned before and is therefore not yet part of the *linguistic* context. In 71 and 72, it would appear that either the noun or the adjective can be accented in both English and Serbian, though Šuštaršič (2005: 43) insists that it is the adjective that attracts the nucleus in Slovene.

In 73, the Serbian participant produced three nuclei as against the two I produced in my performance of the English utterance, thus also accenting *veš* a second time.

Subsection 3.3.10 shows that verbs expressing opinion + *so* or *not* in English attract the nucleus, but in Serbian, if there is a subordinate clause attached to them, then the nucleus is assigned to that clause (74–77).

Comparative constructions (80) appear not to front the nucleus onto the comparative adjective in Serbian. This is the usual pattern in English, too, though Slovene has a fronted nucleus in these cases (Šuštaršič 2005: 42).

3.3.1. Serbian "to be" required in second position unless strong

34. He's come. *Stigao je.*
35. He has come. *Jeste stigao.*
36. That's possible. *Moguće je.*
37. That's not possible. *Nemoguće je.*
38. He's done it. *Uradio je to.*

3.3.2. Serbian topicalization

39. We had chicken for lunch. *Piletinu smo ručali. (Ručali smo piletinu.)*
40. I'm carrying a gift for a friend. *Za druga nosim poklon.*

3.3.3. Event sentences with the nucleus on the subject in English

41. A criminal's escaped. *Kriminalac je pobegao.*
42. The chimney's falling off. *Dimnjak će pasti.*
43. A plane's crashed. *Srušio se avion.*
44. It's raining. *Kiša pada / Pada kiša.*
45. My hammer's broken. *Čekić mi se slomio.*

3.3.4. Names of streets, squares and stations, etc. and related compounds

46. Brankova Street. *Brankova ulica.*
47. Pionirski Park. *Pionirski park.*
48. Hyde Park. *Hajd park.*
49. Bus station. *Autobuska stanica.*
50. Taxi rank. *Taksi stanica.*
51. Savski Square. *Savski trg.*
52. Beovoz Station. *Stanica Beovoza.*

3.3.5. Strong pronouns after prepositions in Serbian

53. Do it for me. / Do it for me. *Učini to za mene.*
54. Tell me about it. *Pričaj mi o tome.*

3.3.6. Narrow focus through anaphora

55. You always have been a sensible person. *Uvek si bio razuman. (… jesi …)*

56. A: Do you like dogs? B: Oh, I like all animals. *A: Da li voliš pse? B: A, ja volim sve životinje.*
57. A: How about a gin and tonic? B: Oh, I'd prefer a vodka and tonic. *A: Da li želiš džin i tonik? B: Uh, ja bih radije votku i tonik.*
58. Do you mind cats? I adore cats. *Da li ti smetaju mačke? Ja obožavam mačke.*
59. Lecturer to guest lecturer: Most of my students are females. *Većina mojih studenata su devojke.* Guest lecturer: Good, I like girls! *Lepo, ja volim devojke.*
60. A: Who brought the champagne? B: Dick brought the champagne. *A: Ko je doneo šampanjac? B: Dik je doneo šampanjac?*
61. A: You took my purse. B: I didn't take your purse. *A: Uzeo si moju tašnu. B: Ja nisam uzeo tvoju tašnu.*

3.3.7. Narrow focus. Emphatic pronoun

62. I saw it. *Ja sam ga video.*
63. Don't they like it? *Zar im se to ne sviđa?*
64. Why can't we go? *Zašto mi ne možemo da idemo?*
65. You know what I think. Now, what do you think? *Znaš šta ja mislim. Ali, šta ti misliš?*
66. I don't know what you're complaining about. *Ne znam zbog čega se ti žališ.*

3.3.8. Narrow focus. Emphatic possessive adjective

67. In my opinion ... *Po mom mišljenju ...*
68. From my point of view ... *Sa moje tačke gledišta ...*
69. Let's go back to my place. *Hajdemo nazad do mog stana.*

3.3.9. Narrow focus not necessarily used in Serbian or English

70. What a lovely day! *Kakav divan dan!*
71. What a nice lady! / What a nice lady! *Kakva divna dama! / Kakva divna dama!*
72. China is a huge country. / China is a huge country) *Kina je ogromna zemlja. / Kina je ogromna zemlja.*
73. A: Shall we wash the clothes. B: Oh, I hate doing the laundry. *A: Da li da operemo veš? B: Uh, ja mrzim da perem veš.*

3.3.10. *Hope*, *think*, *suppose* + *so/not* attract nuclear stress in English, but their equivalents in Serbian do not necessarily do so

74. I think so. *Mislim da hoće.*
75. I don't think so. *Mislim da neće.*

76. I suppose so. *Mislim da jeste (Verujem).*
77. I don't suppose so. *Mislim da nije (Ne verujem).*
78. I hope so. *Nadam se.*
79. I hope not. *Ne nadam se.*

3.3.11. Comparatives

80. Your pronunciation is better than mine. *Tvoj izgovor je bolji od mog.*

3.4. Intensifiers and frequency adverbs

Intensifiers, while receiving nuclear stress in Slovene (Šuštaršič 2005: 36–38, 42), do not appear necessarily to be given the same treatment in Serbian, except in cases of special emphasis, as in 84 and 85. The same is true of frequency adverbs (cf. Šuštaršič 2005: 37–38, 42), as can be seen in 86 and 87.

81. I'm too tired. *Previše sam umoran.*
82. The film was terribly boring. *Film je bio užasno dosadan.*
83. The train is terribly slow. *Voz je neverovatno spor.*
84. A: I want to speak to the manager. B: Mr. Harris is much too busy. *A: Želim da razgovaram sa menadžerom. B: Gospodin Haris je previše zauzet.*
85. A: It's hot! B: You can say that again! *A: Vruće je! B: Zaista je vruće.*
86. He always sleeps; he never works. *Uvek spava; nikad ne radi.*
87. He sometimes comes to class. *Ponekad dolazi na nastavu.*

3.5. Accented prepositions in Serbian

Sometimes prepositions are accented in Serbian, rather than the following noun or pronoun (20–21). This may occur when they are followed by a noun with a falling tune on the first syllable, as in 88 (Hammond 2005: 32). The retraction of the accent onto clitics in general (not only prepositions) is more widespread in western Neoštokavian regions than in the north-east, but instances are found all over the territory alongside cases in which the accent is placed by (more educated) speakers on the more accentogenic accompanying lexical words (Lehiste & Ivić 1986: 171–172).

88. They went towards the house. *Išli su ka kući.*

3.6. Different sentence element selected for focus in Serbian and English

The examples in this section are more difficult to categorize and some, if not all, warrant an individual explanation.

In 89, the intonational pattern in English, with the nucleus right at the beginning of the utterance, on *now*, has a highly pragmatic interpretation expressing exasperation (cf. 104). In the Serbian version it is the verb that receives the nucleus.

In the Serbian version of 90, *još* 'more, else' could be the nucleus but *nešto* 'something' takes precedence. *Nešto* is also accented in 91, whereas in English it is treated as "empty" in this example. Example 92 also has an "empty" indefinite pronoun in English; this time, Serbian accents the verb, which is semantically the most important word in this utterance. Sentence 93 also has the nucleus on the verb in Serbian, as in English, and the word *someone* once again is relatively "empty". In 94, Serbian may represent "empty" *someone* as the noun *drug* 'friend', which can receive the nucleus. In 95, English assigns narrow focus to *thing*, whereas Serbian prioritizes the verb (cf. 89, 93).

In 96, whereas English uses contrastive focus on the pronouns, Serbian accents the noun object in the first clause, and then the pronominal subject in the second.

In 97, since weak auxiliaries cannot be accented in Serbian, the adjective is focussed and given more importance than the verb, too. Examples 98 and 99 also reveal that such auxiliaries are not accented in Serbian, and that past participles often receive the nuclear stress instead.

Sentence 100, while showing that Serbian can accent frequency adverbs for special emphasis (see also 84, 85), also illustrates that, whereas catenative verbs can be accented in English (e.g. *to try*), this is not usual in Serbian.

Sentence 101 exemplifies the fact that final adverbs of place and time are not accented in English unless in narrow focus, whereas in Serbian they are.

The adverb *too* is typically stressed in English, but *i* 'and, too' is not stressed in Serbian, as can be seen in 102, where it is the pronoun *ja* 'I' that becomes the nucleus in the second clause.

The Serbian relative pronoun may take precedence over the verb regarding nuclear stress placement, as can be seen in 103.

Number 104 is idiomatic in English with a characteristic idiosyncratic intonation pattern (cf. 89).

In 105, the effect of anaphora is evident in the English version, while Serbian simply uses a different word order.

The English versions of 106–108 contain redundant elements, which may be omitted in Serbian, so the nucleus goes on the same lexical item in both languages.

Example 109 shows compound stress in English, a pattern that Serbian does not use and therfore accents *poklon* 'present'.

Example 110 shows stress shift to the contrastive adjectival suffix *-ish* in English, a pattern not followed by Serbian.

89. Now what's the idiot done? *Šta je taj idiot sad uradio?*
90. Do you need anything else? *Da li ti je potrebno još nešto? (Cf. Šta ti je još potrebno?)*
91. His name was Billy, or Jimmy, or something. *On se zove Bili, ili Džimi, ili tako nešto.*
92. Would you like a drink of anything? / Can I get you anything? *Da li želiš piće?*
93. I thought I heard someone. *Učinilo mi se da sam čuo nekog.* ("Empty" *someone*)
94. Stop pestering me! Ask Danny or someone. *Prestani da me gnjaviš. Pitaj Denija ili nekog drugog.*
95. She just sat there and didn't say a thing. *Ona je samo sedela tamo, i nije ništa rekla.*
96. I'll make a donation if you do. *Ja ću dati donaciju ako i ti to uradiš.*
97. My! You have done well! *O, pa ti si odlično uradio!*
98. You were good, weren't you. *Ti si bila dobra, zar ne?*
99. A: Are you religious. B: I used to be religious. *A: Da li si religiozan? B: Bio sam religiozan.*
100. A: She's got fat again. B: Well, she was trying to lose weight. *A: Ona se ponovo ugojila. B: Pa, ona je pokušavala da izgubi.*
101. Are you coming to the dinner on Friday? *Da li dolaziš na ručak u petak?*
102. Dave's in the choir, and I'm singing, too. *Dejv je u horu, ali i ja pevam.*
103. I'm going to the party, but I haven't got anyone to go with, though. *Idem na žurku, ali nemam sa kim da idem tamo.*
104. That's a good one! *Ta ti valja!*
105. Football results: Arsenal two, Fulham two. *Arsenal, Fulam 2:2.*
106. Put it on the table there. *Stavi ga na sto.*
107. Chocolate, anyone? *Da li neko želi čokoladu?*
108. There's a man at the door. *Neki čovek je na vratima.*
109. I got her a birthday present. *Kupio sam joj rođendanski poklon.*
110. A: Was it red? B: Well, reddish. *A: Je li bilo crveno? B: Pa, crvenkasto.*

4. Conclusion

Obviously, these brief observations on the position of the nucleus in English and Serbian only draw attention to a few of the most striking differences. Moreover, many of the sentences used are simplex and only contain one intonation phrase. There are other studies which, while not dealing specifically with the nucleus, do offer much more detail. Predolac (2011), for example, is a detailed account of the effect of "constituent order variation" alongside "flexible relative prominence", although he insists that they operate on the sentence independently to produce combined results. Moreover, he investigates the case of sentences with more than one nucleus or which contain bipartite noun phrases (*Srebrne nosim minđuse* 'Silver wear I ear-rings') (Predolac 2011: 148–152). Naturally, different languages obey different word-order systems and, in the words of Firbas (2006: 139), "… language is capable of approaching the extralinguistic reality from different angles and viewing it in different perspectives". Nevertheless, I think that I have been able to point to the main areas of contrast between English and Serbian nuclear stress, showing that English is restricted by a more rigid syntax that has important repercussions for the position of the nucleus and, hopefully, I have provided food for thought and a basis for further studies in this complex area of language.

References

Bolinger, D. 1965. Pitch Accent and Sentence Rhythm. In *Forms of English: Accent, Morpheme, Order*, edited by I. Abe and T. Kanekiyo, 139–180. Cambridge: Harvard UP.

—. 1972. Accent is predictable (if you are a mind reader). *Language* 48: 633–644.

—. 1982. Intonation and its parts. *Language* 58: 505–533.

—. 1986. *Intonation and its parts*. London: Edward Arnold.

—. 1989. *Intonation and its Uses*. Stanford: Stanford UP.

Cruttenden, A. 1990. Nucleus placement and three classes of exception. In *Studies in the Pronunciation of English*, edited by S. Ramsaran, 9–18. London: Routledge.

—. 1997. *Intonation*. Second edition. Cambridge: CUP.

Firbas, J. 2006 (1992). *Functional sentence perspective in written and spoken communication*. Cambridge: CUP.

Hammond, L. 2005. *Serbian. An Essential Grammar*. London: Routledge.

Ladd, R. 1980. *The Structure of Intonational Meaning*. Bloomington: Indiana UP.

—. 2008. *Intonational Phonology*. Second edition. Cambridge: CUP.

Lehiste, I. & P. Ivić. 1986. *Word and Sentence Prosody in Serbocroatian*. Cambridge, MA: Massachusetts Institute of Technology.

Mott, B. 1993. The intonation of English and Spanish: contrastive analysis. In *Actas del XV Congreso de AEDEAN (Asociación Española de Estudios Anglonorteamericanos)*, edited by F. J. Ruiz de Mendoza and C. Cunchillos, 621–632. Logroño: Colegio Universitario de la Rioja.

—. 2009. Fronted nuclei: narrow focus in Spanish and English. *Bucharest Working Papers in Linguistics*, vol. XI, nr. 2, 73–82. Bucharest: University of Bucharest.

—. 2011. *English Phonetics and Phonology for Spanish Speakers*. Segunda edición. Barcelona: Publicacions i Edicions de la Universitat de Barcelona. (Especially ch. 10, *Intonation*)

Predolac, N. 2011. *Syntax and information structure: free constituent order and flexible relative prominence in Serbian*. Unpublished PhD dissertation, Cornell University.

Selkirk, E. 1984. *Phonology and syntax. the relation between sound and structure*. Cambridge, Ma.: MIT Press.

—. 1995. Sentence prosody: Intonation, stress, and phrasing. In *Handbook of phonological theory*, edited by J. A. Goldsmith, 550–569. London: Blackwell.

—. 2002. Contrastive FOCUS vs. presentational focus: Prosodic evidence from right node raising in English. In *Speech Prosody 2002: Proceedings of the 1st International Conference on Speech Prosody*, edited by B. Bel and I. Marlien, 643–646. Aix-en-Provence: LPL – CNRS/SproSig.

Šuštaršič, R. 2005. *English-Slovene contrastive phonetic and phonemic analysis and its application in teaching English phonetics and phonology*. Ljubljana: Znanstveni inštitut Filozofske fakultete. (Especially section 3: Accentuation and position of the nucleus in English and Slovene, 33–44.)

Wells, J. C. 2006. *English Intonation. An introduction*. Cambridge: CUP.

METHODOLOGICAL ISSUES IN THE ACOUSTIC ANALYSIS OF SPONTANEOUS SPEECH PROSODY

ALEKSANDAR PEJČIĆ

Outline

In this chapter, the author presents the methodological design of a research study which aimed to investigate the prosodic characteristics of Serbian and British persuasive political speech, and the methodological issues he encountered in the process. These issues can mostly be related to the choice of suitable speech tokens, in terms of their subject, register, style, as well as the regional, gender and age differences of the speakers. Other issues contributed to the final design of the questionnaire used in the survey, while the variety of sources used for the extraction of speech tokens also raised some questions. In addition, the acoustic analysis itself demanded that the author address any errors and subsequent repairs made by individual speakers. The author illustrates his points by the data obtained in this research study and points to possible pedagogical implications of spontaneous speech analysis.

1. Introduction

While both recorded and spontaneous speech provide different benefits for the researcher, spontaneous speech has been comparatively less popular in acoustic speech research since laboratory conditions and the use of read speech offer more control over the analyzed content, while reducing the possibility of error. However, in cases when only spontaneous speech will fit the research subject, additional effort must be invested into ensuring that the results should not be compromised by the methodological issues involved in such a study.

The aim of this paper is to present the methodological issues stemming from the use of spontaneous speech, which became apparent in the course of the research undertaken by the author in his MA thesis, titled *What do We Believe? Prosodic Correlates of Persuasive Speech in Serbian and English Political Discourse,* as well as the measures taken in order to minimize the effects of these issues on research results produced at various stages of the analysis. The potential applications of findings resulting from spontaneous speech analysis will also be mentioned.

2. Theoretical background

One of the issues addressed in this paper, namely, the difference between read and spontaneous speech, is usually referred to as the difference between speaking styles.

Being mostly preferred because of its suitability for controlled, laboratory conditions ideal for eliciting speech, read speech has been the favoured option among phoneticians dealing with both segmental and supra-segmental units of speech. The controlled environment also means that the researcher is directly involved in speech production and able to alter and adapt the parameters of the experiment and any elicited material to suit immediate needs, or correct any errors in the experimental process. This is especially important in the elicitation and analysis of emotional content in speech, where the unwritten rule is the use of actors for the task of displaying extreme emotions[1] (Athanaselis et al. 2005). Read speech also gives the researcher more control over inter-speaker and intra-speaker differences (age, gender, dialect, native/non-native speakers, physiological and psychological conditions…), while spontaneous speech often consists of pre-recorded material where such variations are not only more difficult to control, but at times impossible to pinpoint. Within the past decade, read speech has been given priority over spontaneous speech in the development of Automatic Speech Recognition (ASR) software, mainly due to its high recognition accuracy (Athanaselis et al. 2005). Finally, it should be mentioned that the long tradition of research based on read speech has established not only plenty of literature on its use, but a significant number of speech corpora as well.

[1] Usually fear, anger, joy and sadness.

Certain remarks, however, can be made about the use of read speech. Firstly, it seems that emotional speech analysis, while so far mainly relying on the 'acting out' of scripts by professional actors, does not cover the full range of emotions, nor does it deal with complex emotions expressed in everyday speech (Athanaselis et al. 2005). Even though it is difficult to pinpoint specific emotions in spontaneous speech (Scherer 2003), present knowledge about read emotional speech should be a good basis for such advancement. Problems also exist for the recognition of everyday speech by ASR software, while an additional point can be made about the possible lack of adequate laboratory conditions for the elicitation of read speech, or their lack of availability to young researchers.

Spontaneous speech, however, defined by Beckman (1997, cited in Hansson 2003: 25) as "speech that is not read to script", is widely available and can either be found on various public domains or be elicited by means of different techniques, which will be revisited later. Unfortunately, spontaneous speech gives the researcher little control over content and syntactic structure (Hansson 2003), which can sometimes be broken by background noise and interruptions, especially when the speech material has been recorded in a TV studio environment. To no lesser extent, spontaneous speech suffers from sloppy pronunciation, disfluencies, repairs, filled pauses, elongated segments, editing expressions (e.g. *I mean, you know*), word fragments, and repeated words (Nakamura et al. 2008; Clark & Wasow 1998), all of which can hinder acoustic analysis, especially on the supra-segmental level. Finally, we should mention that even in the field of ASR software development, read speech has achieved far higher rates of recognition than spontaneous speech, although the linguistic and phonetic models used in this type of research have been, in fact, devised primarily for read speech (Nakamura et al. 2008).

On the other hand, spontaneous speech can be more useful in conversation analysis since conversations and narratives are more frequent than read speech and "social interaction is constantly constructed and made sense of by the interactants themselves, rather than given *a priori*" (Szczepek Reed 2006: 22). Furthermore, the relationship between prosody and pragmatics is best observed by means of spontaneously constructed expressions (Braga & Marques 2004) and can provide a better understanding of prosody and discourse organization than read speech (Beckman & Venditti 2000). In addition, the recent creation of spontaneous speech corpora (Santa Barbara, TED, SWITCHBOARD, CSJ, etc.) makes various recorded material (both elicited and pre-

recorded) accessible for a variety of potential uses in acoustic and other linguistic research.

This material, according to Llisterri (1992: 2–3), can be classified into 3 types and 2 subtypes, based on a sample of 15 papers presented at the ESCA Workshop on Speaking Styles, (Barcelona 1991):

1. Samples recorded in laboratory conditions
   1.a with non-professional speakers
   1.b with professional speakers
2. Samples obtained from recorded TV or radio broadcasts
3. Samples recorded in the speaker's natural environment

Here, individual labels for speech material belonging to these speaking styles range from a free conversation with a friend (1a), political debates and newscasts (2), to interviews with native speakers (3), with many other examples supplied by Llisterri (1992: 2–3). Naturally, along with the proliferation of Internet technology and social networking tools, the number of possible sources for speech material has also increased, and now includes sources such as voice chat platforms, video blogs, podcasts and other web-based content.

However, while Llisterri focuses mainly on the type of subject and setting used for the production of spontaneous speech material, Beckmann (1997, cited in Hansson 2003: 26) lists 10 techniques for its elicitation by the researcher: the unstructured narrative, extended descriptive narrative, instruction monologue, instruction dialogue, database querying dialogue, Wizard of Oz, performance narrative, overheard conversation, enacted conversation and the public conversation technique. In addition, Hansson (2003) states that in the course of a phonetic research based on spontaneous speech, the researcher should go through the process of first choosing the elicitation technique, a sufficiently good recording and a communicative situation "allowing a reasonable degree of control over linguistic content and discourse structure" (Hansson 2003: 26) in order to ensure the quality of final results.

In the following section, we will present the outline of the methodological process of choosing suitable tokens for acoustic analysis, explain how it follows the three points made by Hansson (2003) and expound on the effects it had on the final acoustic analysis performed in the MA thesis.

3. Present study

Right away, we must draw attention to the fact that no elicitation technique was used for the production of spontaneous speech in this research. In order to get the spontaneous political discourse we needed for our analysis, the material used in this study contained what Llisterri (1992) would refer to as 'samples obtained from recorded TV or radio broadcasts', consisting of 10 speech samples extracted from speeches of 5 Serbian and 5 British politicians taking part in different televised debates in Serbia and Great Britain. To make sure the presented material was of sufficiently good recording quality, while containing topics relevant to the listeners, Serbian speech samples were taken from recent shows aired on RTS 1 (*Da Mozda Ne* and *Upitnik)* and B92 (*Utisak Nedelje)* TV stations, made publicly available on their websites and dating only as far as April 2011, while the English samples were taken from the BBC's election debate taking place in April 2010 and the BBC's show *Question Time* from October 2009, made public on and downloaded from Youtube.

The following sections will address Hansson's (2003) third point, namely that the communicative situation present in the speech material should give the researcher a satisfactory amount of control over spoken content. In the present research, control over the communicative situation was achieved in several steps, where the issues addressed were:

- Setting (interview, panel show, discussion, presentation…)
- Topic (domestic/foreign, social, political, economic…)
- Inter-speaker variation (dialect, gender, age)

3.1. Setting

Regarding the setting in which the recorded speeches took place, the televised political debate was chosen, partly because of the existing literature on the subject (Braga & Marques 2004, Rosenberg & Hirschberg 2005, 2009, Touati 1991), and partly because it gave a relatively controlled sound quality, having been recorded in a studio environment. That meant it would also entail less background noise and fewer interruptions than a recording of a parliamentary session. There was also no turn taking and speakers were usually allowed to form longer narratives, often as a response to a host's question or another participant's statement or remark.

3.2. Topic

The issue of topic was important because, depending on the subject of the debate, more or less heated responses from participants could be expected, and a controversial or important topic would serve as an ideal vehicle for differing opinions. Our Serbian speakers discussed domestic election issues and their government's social policies, and both government and opposition speakers were all well-known politicians, including: Vuk Drašković (Serbian Renewal Movement), Tomislav Nikolić (Serbian Progressive Party), Nenad Čanak (League of Vojvodina's Social-Democrats), Rasim Ljajić (Social Democratic Party of Serbia) and Velimir Ilić (New Serbia). The chosen British speakers discussed the British immigration policy (an important issue at the time of the election, and the one most likely to get the Serbian survey participants interested in the subject and the content of the speeches), and again both government and opposition speakers were well-known British politicians: Gordon Brown and Jack Straw (Labour Party), David Cameron (Conservative Party) and Nick Clegg and Chris Huhne (Liberal Democrats).

All speakers clearly expressed different standpoints and attitudes towards a sensitive topic, and did so with a distinctive speaking style, characterized by perceptibly different uses of prosodic cues (most importantly the prosodic properties of pitch, loudness and speaking rate). Such a selection of prosodically distinctive speaking styles was made possible by the fact that we opted for the use of the more popular and well-known politicians. Had less well-known speakers been used, especially in the case of Serbian politicians, this would have been difficult to achieve, and the prosodic differences in their speeches would not have been clear enough to catch the attention of the listeners evaluating their persuasiveness, causing other linguistic and extra-linguistic factors to become more prominent than the prosodic ones we set out to analyze.

3.3. Inter-speaker variation

Dialect, gender and age differences were also important to control, as their effects on different prosodic cues have been well documented. Therefore, all five Serbian speakers we chose were born in Serbia (two in Vojvodina, three in Central Serbia), in regions where standard Serbian dialects are spoken. Likewise, all five English speakers we chose were born in Great Britain (four in England, one in Scotland), in order to avoid prosodic variation between British and American English. Regional

differences between British speakers could not be avoided, although we are aware of the differences in F0 range, F0 modulation, tempo, etc. noted in the acoustic analyses of the said variants. Because gender-based variation is known to affect different pitch characteristics, all ten chosen speakers were males.

The age of the speakers ranged from 43 to 63 (British) and from 47 to 64 (Serbian), while the mean age was 52.8 (st. deviation 8.96) for the British speakers, and 56.4 (st. deviation 6.09) for the five Serbian speakers. In this particular case, the effects of age on prosody such as a decreasing lung capacity, calcification of the larynx, loss of muscular control causing a decrease in vocal intensity, lowering of vocal pitch, and narrowing of the vocal pitch range (Amilon et al. 2007) were not found, as some of the most experienced politicians in our experiment exhibited prosodic characteristics which did not match Amilon's findings.

By taking notice of these issues, we avoided making an inappropriate or irrelevant selection of spontaneous speech samples which could have made our data analysis much more problematic. However, these issues were mostly speaker and context-oriented, and for the most part indirectly related to the acoustic analysis.

With the selection of speakers and tokens made, the following steps in our research dealt with the preparation of tokens for the speech analysis, which was done using Praat (Boersma & Weenink 2010). In order to ensure the integrity of prosodic material, which was our primary interest at this point, the following additional issues had to be dealt with:

- Token duration and content
- Sound quality and normalization
- Speech extraction and segmentation
- Speech disfluencies and pronunciation issues

3.4. Token duration and content

Regarding token duration and content, what we set out to achieve was a balance between token length, the number of tokens to be played to a sample of listeners, and a suitable survey duration of 25 minutes, during which the speech tokens were to be played and a questionnaire about their content filled out by the participants. In order to achieve this, we opted for 10 spontaneous speech samples (5 Serbian, 5 British), with a duration range of 21s to 30s, and a mean duration of 25.9s. This, we estimated, was

long enough for the speakers to express an entire thought or message and for the participants to form an opinion about the speaker and speech, but brief enough to keep the participants' attention. This had to be achieved with two things in mind. Firstly, the selected tokens had to have clearly marked prosodic and syntactic boundaries, often separating the selected tokens from interruptions occurring during the debates. Secondly, spontaneous speech is often characterized by a higher speech rate, and a fast speech with few pauses can make the process of marking token boundaries more delicate, which was the case even with a few speakers in our study.

3.5. Sound quality and normalization

Even though political debates recorded in TV studios have relatively stable recording conditions, there are many interruptions (host, other speakers, audience), and different speech token sources mean different recording conditions (microphone distance, audio settings, studio acoustics, etc.). Under such circumstances, normalization of audio material is usually undertaken to ensure equal loudness for tokens coming from different sources. However, different software offers different normalization tools, and instruction on how to normalize speech tokens for acoustic analysis is difficult to find. An example by Rosenberg & Hirschberg (2009), who equalized their tokens for loudness to –12db and only analyzed mean and standard deviation figures of intensity after normalization, shows that normalizing audio material has both advantages and disadvantages. On the one hand, it reduces the number of operations available in the acoustic analysis, while on the other, it improves the methodology of the research. In this study, having faced several problems mentioned in this paragraph, we decided not to normalize the speech tokens for intensity, but rely instead on the intensity variables that are unaffected by normalization, i.e. mean intensity and the standard deviation of intensity.

3.6. Speech extraction and segmentation

Once the issue of normalization was cleared, the ten selected speech tokens could be prepared for analysis in Praat (Boersma & Weenink, 2010), where the fundamental frequency and intensity were set to be computed and shown in the following manner:

- F0 (Hz): 50–300 Hz
  cross-correlation analysis method
  automatic drawing method
- Intensity: 50–100 dB
  Mean dB averaging method
  Subtract mean pressure

These lowered floor settings for pitch analysis allowed us to analyze male voices of lower fundamental frequency, including one creaky male voice, while covering all pitch peaks with the standard ceiling settings. Regarding intensity, the settings remained at standard values, with the mean pressure subtraction on, given the variety of recording conditions. In addition, the threshold for the analysis of pauses was set to 200ms, although differing opinions exist on this matter, as will be explained in the following paragraph.

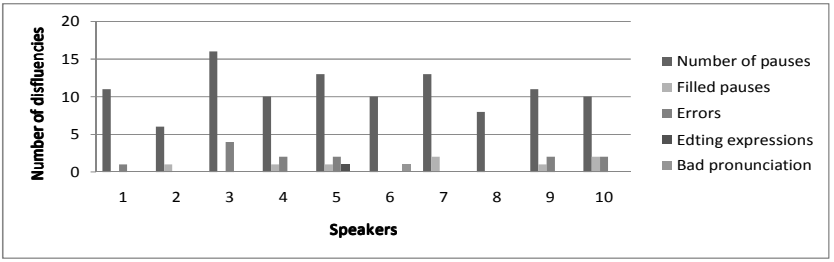
On the one hand, most studies have taken a silent period of over 200ms or 250ms based on a practice first introduced by Goldman-Eisler (1968) (as mentioned in Zellner 1994), but on the other, lower thresholds (0.1s) can be found in Wennerstrom & Siegel (2003) while Hieke, Kowal & O'Connell (1983) have supplied evidence that the threshold of 250ms cannot always successfully separate articulation from hesitation pauses. Furthermore, the practice of using thresholds itself was later criticized by Campione & Veronis (2002) who felt that using thresholds can severely impair research results and opted instead for a classification into brief (< 200ms), medium (200–1000ms) and long pauses (> 1000ms), where the latter can be found only in spontaneous speech (Campione & Veronis 2002: 199). In our study, we chose to rely on the practice established by the most influential research on pauses, although the point made by Campione & Veronis may be pursued in the following papers.

3.7. Speech disfluencies and pronunciation issues

As our research progressed and we reached the stage of acoustic analysis, the final issue of speech disfluencies and pronunciation issues came to light. It is well known that spontaneous speech suffers from sloppy pronunciation and disfluencies including repairs, filled pauses, elongated segments, editing expressions (e.g. I mean, you know), word fragments, and repeated words (Nakamura et al 2008, Clark & Wasow 1998). Pauses, as well as other disfluencies in speech, are an inevitable part of spontaneous speech and can appear out of hesitation and insecurity,

or signal "cognitive planning processes" (Butterworth 1975: 75), whereby the speaker gives himself additional time to plan what he is about to say, or to 'retrace his steps' once he has made an utterance and realized that his speech "has raced ahead of cognitive activity" (Zellner 1994: 47). Whatever their cause is, their role in spontaneous speech analysis deserves special attention (Chart 1).

Chart 1. Number of disfluencies found per speaker



Unfortunately, some of the disfluencies we found could not be tolerated, as they disrupted the syntactic and prosodic units we needed for the acoustic analysis. These errors and repairs changed the intonation contours of sentences or the type of focus and its placement within separate intonation units, which led to our decision to remove these fragments from the acoustic analysis in order to prevent our own error in their classification and consequent description. Since we were only analyzing the four most and least persuasive speakers at this point, only two speakers (Speaker 2 and Speaker 10, both Serbian) were found to have made such errors. However, the possibility of finding more of such fragmented prosodic units in spontaneous speech samples is high, and a systematic filtering out of such errors during their acoustic analyses might be advised, especially for inexperienced researchers.

So far, we have mentioned the issue of filled and silent pauses and certain major and minor disfluencies in this research. While the former were not unwelcome in this study, since pauses were, alongside pitch, intensity and speech rate, one of the analyzed prosodic characteristics of speech, the latter had to be addressed in terms of being either tolerated or filtered out, depending on their effect on the prosodic content. In addition, various pronunciation issues created problems in the annotation of

prosodic cues, especially pitch and intensity. These issues can be seen in the list below and include:

- Devoiced boundary tones and the loss of F0 contour (*stranaka, izborima, permits…*)
- Breaks in the F0 signal (due to devoicing and voice quality)
- Blended words due to fast speech (*petnešesnes* instead of *petnaest šesnaest)*
- Swallowed syllables containing plosives (*nepularno* instead of *nepopularno)*
- Shortened unaccented vowels and syllable elision

As with the disfluency issues, the majority of such problems did not pose a serious threat to our research results, and can actually be expected in any speech, especially a spontaneous one. Even though our speakers were experienced politicians and could be expected to have full control of their diction and articulation, several speakers, Serbian in particular, had difficulty enunciating certain types of words, mainly as a result of faster speech. In such cases of blended words, swallowed syllables and syllable elision, the issue was primarily on the segmental level with which we were not concerned, and intonation cues (including pitch accents and boundary tones) were observed relative to the way the words and phrases were uttered.

Slightly more complicated was the situation with the breaks in the F0 signal, either at word boundaries as a result of devoicing, or in the middle of words, related mainly to the voice quality of one particular speaker. The issue here existed mainly because of the fact that most of our acoustic analysis was quantitative and the information on pitch movement we were interested in was missing at times, which made the correct annotation difficult. However, as with the previous issue, we could only work with the signal that was recognized by the speech analysis software, and the annotation of acoustic cues was made at points available in the acoustic image shown in Praat.

Finally, having dealt with the issues discussed above, the acoustic analysis of pitch, intensity, speech rate and pauses in political speech samples could be performed. It was done in three steps, starting with all 10 speech samples being analyzed in terms of their global properties. The second stage consisted of a descriptive analysis of the four most and least persuasive speeches, divided into broad and narrow-focus intonation units (Baumann et al. 2006), while in the third a comparative approach was used on the same number of samples. The list below shows the prosodic cues analyzed in the course of this analysis.

Pitch

- Max, min, range, mean and the standard deviation in entire tokens
- Max and range in the first stressed syllable in the broad focus domain
- Max and range in the first nuclear syllable
- Type of nuclear tones
- F0 of the IU boundary tone
- F0 min, max and range of entire intonation units
- Duration of nuclear syllables

Intensity

- Mean and standard deviation of entire tokens
- Mean and standard deviation of focus domains and background
- Max and range on first syllables of broad focus domains and words carrying nuclear stress (only for individual tokens)

Speech rate

- Speaking rate (speech rate with pauses)
- Articulation rate (speech rate without pauses)
- Speaking rate within separate intonation units

Pauses

- Silent and filled pauses
- Number, duration, average duration and percentage in speech tokens
- Pause placement and type

4. Conclusion

To sum up, this paper offers an insight into the process of preparing spontaneous speech material for a successful acoustic analysis of its prosodic structure. This includes the selection of suitable material for the analysis, its preparation for the operations to be performed using speech analysis software, and the handling of various errors in the prosodic content. One can argue that in a case like this, the costs of using spontaneous speech for acoustic analysis outweigh its benefits. Certainly, differences in the recording environment and the prosodic and syntactic content of chosen speech tokens would not be as big an issue had read speech been used in this particular study. On the other hand, some of the problems facing the use of spontaneous speech are not foreign even to the more reliable and systematic use of read speech, and include the choice of

speakers, as well as pronunciation and disfluency issues. The choice then, is up to the researcher, and the subject and application of the study for which the acoustic analysis is performed.

Most of the subjects and applications of spontaneous speech research are well known, and include the analysis of the relationship between speech acoustics and linguistic pragmatics, the development of ASR systems, the teaching of foreign language prosody, conversation analysis, etc. In addition, the aim of the MA study to which this paper refers was to establish which prosodic characteristics are related most closely to skilful public speaking, and can be learned through training, and apply these findings to improving class lectures, conference presentations, public speeches, TV appearances, communication and leadership skills. Even a study of a small number of analyzed tokens, and with a specific set of speech characteristics in mind, can offer important information about the use of prosody in spontaneous, everyday speech, and should be encouraged.

References

Amilon, K, J. van de Weijer & S. Schötz. 2007. The impact of visual and auditory cues in age estimation. *Speaker Classification* II: 10–21.

Athanaselis, T., S. Bakamidis, I. Dologlou, R. Cowie, E. Douglas-Cowie, & C. Cox. 2005. ASR for emotional speech: clarifying the issues and enhancing performance. *Neural Networks Elsevier Publications* 18(4), 437–444.

Baumann, S., M. Grice & S. Steindamm. 2006. Prosodic marking of focus domains – Categorical or gradient? *Proceedings of Speech Prosody 2006*, 301–304. Dresden, Germany.

Beckman, M. E. 1997. A typology of spontaneous speech. In *Computing Prosody. Computational Models for Processing Spontaneous Speech,* edited by Y. Sagisaka, N. Campbell & N. Higuchi, 7–26. New York: Springer.

Beckman, M. E. & J. J. Venditti. 2000. Tagging prosody and discourse structure in elicited spontaneous speech. In *Proceedings of the Science and Technology Agency Priority Program Symposium on Spontaneous Speech: Corpus and Processing Technology*, 87–98. Tokyo, Japan.

Boersma, P. and D. Weenink. 2010. Praat: doing phonetics by computer (Version 5.2.03). [Computer program]. Retrieved 20th November 2010 from http://www.praat.org/.

Braga, D. & M. A. Marques. 2004. The pragmatics of prosodic features in the political debate. In *Proceedings of Speech Prosody 2004*, 321–324. Nara, Japan.

Butterworth, B. 1975. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research* 4: 75–87.

Campione, E. & J. Véronis. 2002. A large-scale multilingual study of silent pause duration. In *Proceedings of the Speech Prosody 2002 Conference,* edited by B. Bel and I. Marlien, 199–202. Aix-en-Provence: Laboratoire Parole et Langage.

Clark, H. H. & T. Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37: 201–242.

Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in Spontaneous Speech*. London/New York: Academic Press.

Hansson, P. 2003. *Prosodic Phrasing in Spontaneous Swedish*. Unpublished PhD dissertation, Lund University.

Hieke, A. E., S. Kowal, & D. C. O'Connell. 1983. The trouble with 'articulatory' pauses. *Language and Speech* 26: 203–214.

Llisterri, J. 1992. Speaking styles in speech research. In *Proceedings ELSNET/ ESCA/ SALT Workshop on Integrating Speech and Natural Language.* Dublin, Ireland.

Nakamura, M., K. Iwano & S. Furui. 2008. Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language* 22 (2): 171–184.

Rosenberg, A. & J. Hirschberg. 2005. Acoustic/prosodic and lexical correlates of charismatic speech. Presented at *EUROSPEECH'05*, September 2005.

—. 2009. Charisma perception from text and speech. *Speech Communication* 51: 640–655.

Scherer, K. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40: 227–256.

Szczepek Reed, B. 2006. *Prosodic Orientation in English Conversation*. Basingstoke: Palgrave.

Touati, P. 1991. Temporal profiles and tonal configurations in French political speech. *Working Papers* 38. Department of Linguistics, Lund University, 205–219.

Wennerstrom, A. & A. F. Siegel. 2003. Keeping the floor in multiparty conversations: Intonation, syntax, and pause. *Discourse Processes* 36: 77–107.

Zellner, B. 1994. Pauses and the temporal structure of speech. In *Fundamentals of speech synthesis and speech recognition*, edited by E. Keller, 41–62. Chichester: John Wiley.

THE STATUS OF INTONATION IN A LEVEL APPROACH IN THE ORGANIZATION OF LANGUAGE

VLADIMIR PHILLIPOV

I dedicate this paper to Professor Maya Pencheva for all she taught me during our discussions when we shared an office together.

Outline

Traditionally, intonation is the "wandering Jew" between phonology and syntax, yet is treated as being of little linguistic value. In later times, the generativist tradition hardly touched on it, and only recently, the latter has been suggested to convey postlexical pragmatic meanings, i.e. it occupies the 'safest' component that comes hierarchically after syntax.

The present paper views intonation as an exponent of fluctuation, i.e. a shift in the status of a linguistic item leading to a different function, while preserving the form. The analysis attempts at establishing a correlation between syntax and intonation, sheds light on controversial issues of prosodic phrasing and the intonational meaning/ form interface.

1. Introduction

Ever since the inception of linguistics as a domain of scientific study, intonation has been its Cinderella, and its status within a hierarchical/ multidimensional approach to language has invariably been rather tentative. Traditionally, it has had the unfortunate lot of the 'wandering Jew', circularly travelling from phonology to syntax and back, and stopping to take a breath at either destination depending on the linguistic predilections of its begetters (cf. the Bulgarian Academic Grammar, 1982, 1983 where it is given two methodologically different treatments – one within a phonetic/ phonological framework and another within a syntactic one – in two respectively different volumes). As a special privilege it is

included in the grammar of the specific language, yet it is humbled down to an appendix position (cf. Quirk *et al.* 1985). Serbo-Croatian (or Serbian and Croatian as the language has been divided recently) was extensively described within an instrumental framework in Lehiste and Ivić (1986), but the analysis is certainly in need of updating in view of current advances in the domain.

In another paper (Phillipov 2011), I argued the case for the sign character of intonation in language. It was indirectly hinted that Saussure's notion of *linearity* – the postulate that in language one can say only one thing at a time – virtually made prosodic analysis impossible, hence all subsequent attempts to harness it to the reins of some kind of conceptual modification of segmental analysis. The initial step to avoid such a pitfall would be to establish the place of intonation within core grammar, or even better, within a broader language model.

2. Level models of intonation analysis

The linguist first to approach the role of 'levels' in the systematic analysis of language was the French scholar Émile Benveniste who adumbrated the model in his paper *Les niveaux de l'analyse linguistique* delivered at the 9th International Congress of Linguists, held at Cambridge, Mass., 1962. In it intonation plays hardly any part: "Les variétés d'intonation n'ont pas valeur universelle et restent d'appréciation subjective." (Benveniste 1966: 128). Taking into account the state of the art of intonation studies at the time and the strong theoretical impact of A. Martinet's views on the matter, Benveniste's position concerning the place of intonation in a level model of language comes as no surprise (cf. Martinet 1964).

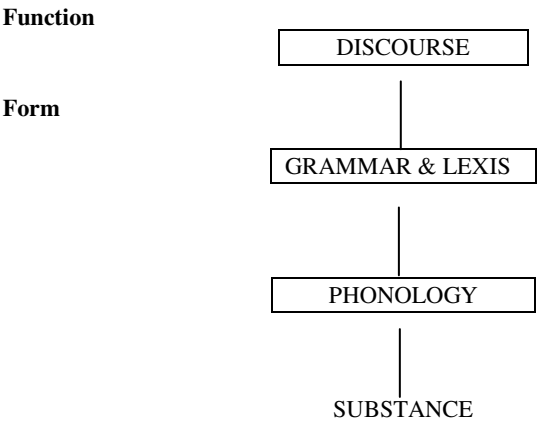
The idea that intonation, the 'wandering Jew' of linguistics, has a special place in the level, hierarchical organization of language I adopted from Professor Maya Pencheva's article on the nature and status of word formation in the system of language (Pencheva 1983). Therein, Pencheva convincingly argues the case for treating word formation as a *suprasystem* which goes through all levels but is not specifically located on any one of them. Its spheres of intensity depend on the typological specificities of the respective language. The prosodic system of language, intonation in particular, can be viewed as an analogous system. In terms of substance intonation is related to phonology, functionally it acts on units of all levels, and communicatively is manifested in speech. Semantics can also be claimed to be a suprasystem, embracing all linguistic levels, but the

latter does that in a qualitatively different way. The systems of intonation and the products of word formation interact with the two-sided language units, i.e. the sign, whereas semantics itself is merely one side of these units, namely the plane of content.

In this line of thought, the three above-mentioned systems are outside the hierarchy of levels, and yet they are systems of a different rank. The basis of the level model is formed by semantics: all levels refer to meaning and at the top of the model is invariably intonation. In this respect, Pencheva (1983: 58) makes a provoking remark: "The bottom and the top constructs of the model ignore the discrete character of language, while the systems in between them are based precisely on the discreteness of the language units."[1]

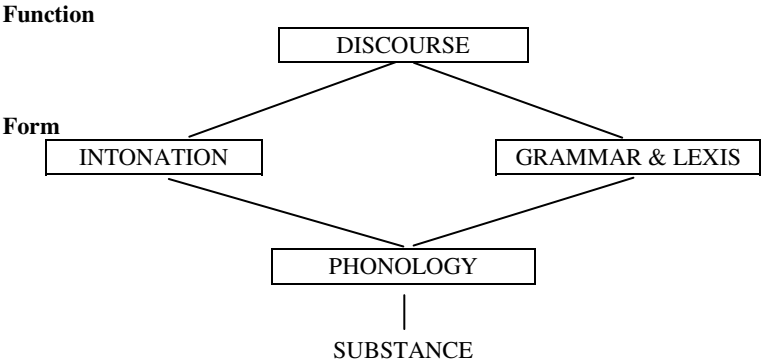
Pencheva's model resembles to a high extent the level configuration proposed by the representatives of the Birmingham School of discourse analysis (cf. Coulthard & Montgomerry 1981). This is a substantial revision of the Hallidayan analysis, the latter going back to Halliday's programmatic article (Halliday 1961) and even further back to the British tradition laid by J. R. Firth. Compare Halliday's diagrammatic representation of linguistic levels, with a discourse level added (Diagram I), with that of the Birmingham School (Diagram II):

Diagram I



[1] The special sign character of intonation has been argued for elsewhere (cf. Phillipov 2011).

Diagram II



As can be seen, both approaches are functionally orientated, the only difference being the positing of a parallel autonomous intonation component in Diagram II. In Halliday's theory (cf. Halliday 1967; Halliday & Greaves 2008), intonation is a department of phonology, and the paradigmatic oppositions it makes available are similar to those of tense, number and mood. Thus, the opposition between what Halliday posits as Tone 1 (falling) and Tone 4 ((rising-) falling-rising) (Halliday 1967: 17) is a realization of several grammatical systems based on the grammatical context. However, as Brazil (1981: 148) rightly points out:

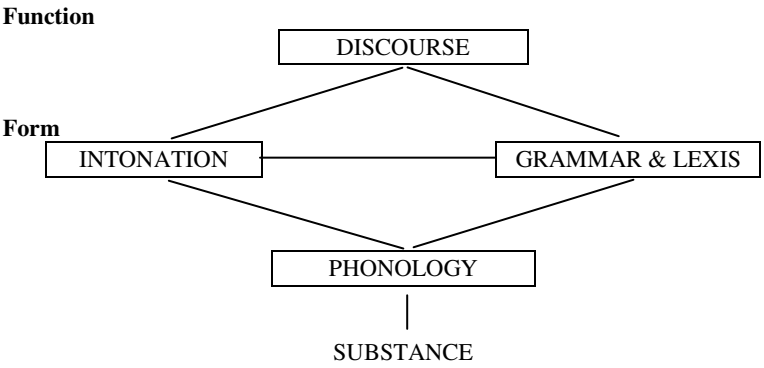
An interesting, though easily overlooked, feature of Halliday's description is that labels like tone 1 and tone 4 are, in effect, both formal and phonological labels. The terms of a system whose realizations are tones 1 and 4 are always realized by the same abstraction from the phonetic data. This contrasts markedly with the way non-intonational systems are realized. The traditional argument for setting up a separate level of phonology has two aspects: plural number, for instance, is not always realized by the phoneme /s/; neither does the phoneme /s/ always realize plural number. The fact that such condition of double articulation does not hold for intonation is one reason for further modifying the model [of Diagram I – VP] to [Diagram II – VP].

Diagram II allows for the postulation of a set of options to which the label of 'intonation' can be applied on a par with 'lexis and grammar'. What is pertinent to the further discussion is that intonation contrasts are syntagmatically conceived as functioning on the same level of abstraction as such relations as 'subject of …' and 'object of…'. Brazil's justification

for setting up the description of 'intonation' and 'lexis and grammar' linking function and substance is based intuitively upon "the observation that there is significantly less opacity in the intonational channel: the theoretical apparatus necessary to deal with the chain that connects sound to sense is less elaborate than that needed to deal with syntax" (Brazil 1981: 149). Analyzing the typological nature of the morpheme in word formation and establishing an explicit analogy with intonation, however, Pencheva (1983: 59) provides deeper insight into the relationship 'substance – form': "Due to meagre morphological marking [=formal marking, VP] the isolated elements [in word formation, VP] are of indefinite nature, and they remain so till they are embedded in a syntagmatic context, even if the latter is the most elementary. *This gives priority to syntax* [italics mine, VP]".

In the generativist tradition, intonation was usually either rejected as being part of competence or it was included in grammar as part of the phonological component linked with stress phenomena. Selkirk (1986) was the pioneer to argue that that prosody – and by extension – intonation must be represented by a set of metrical trees, which are independent of syntax, the former expanding autonomous categories. These categories are to be matched with syntactic trees at surface structure level. However, Ronat (1984) provides evidence that intonation must look at S-structure, at Logical Form and most probably at deep structure, i.e. translated in functional terms there must exist a bidirectional link between 'intonation' and 'grammar and lexis', thus giving as first approximation:

Diagram III



On this model, all components are bidirectionally linked, hence no need of arrows.

Last but not least, the issue of meaning inevitably comes to the fore. On this point Benveniste is unequivocal as to the general principle: "Forme et sens doivent se definer l'un par l'autre et ils doivent ensemble s'articuler dans toute l'étendue de la langue. Leurs rapport nous paraissent impliqués dans la structure même des niveaux et gans celle des fonctions qui y respondent, que nous désignons ici comme "constituent" et "integrant" " (Benveniste 1966: 126). With respect to intonation Halliday and Greaves come up with the hypothesis that is to be validated/ invalidated for specific languages – and it has already been validated for English – namely that "systems of TONE (falling, rising & c.) construe interpersonal meanings, while systems of TONALITY (division into tone units) and TONICITY (location of prominence within the tone unit) construe textual meanings. ***Tone sequences*** (the sequential choices of tone in successive tone units) play some part in construing logical meanings. The only one metafunction to which intonation makes no contribution is the experiential" (Halliday & Greaves 2008: 97).

In this paper, intonation will be analysed on the levels of phonology and syntax. The semantic impact of meaning will duly be taken into account although I shall not touch on the semantic componential features.

3. Intonation interacting with features from segmental phonology

In principle, intonation can 'work' meaningfully together with any linguistic unit, starting from an exponent of a feature(s) of a phoneme, i.e. at the level of phonology, going to the discourse level. In this respect intonation, together with the phonological feature it functions jointly, demonstrates various degrees of fluctuation (cf. Molhova 1981), the latter being defined as a shift in the status of a linguistic unit leading to a different function, while preserving its form. In the process of fluctuating from a bottom-to-top direction intonation leaves traces at every level it traverses. In this line of reasoning, intonation does not merely convey postlexical pragmatic meaning in a linguistically structured way as the generativists would claim it (cf. Ladd 2008); its substantial characteristics lie in phonology acting as what Blake refers to as a "competing mechanism" (Blake 1994: 13) but its formal exponents function explicitly on all the other linguistic levels. "Le *sens* en effet la condition

fondamentalle que doit remplir toute unite de tout niveau pour obtenir statut linguistique." (Benveniste 1966: 122). No matter what level its dominant semantic features operate on, it is a Janus-like creature and influences concomitantly all the other levels and sub-levels, too.

A phonologically motivated emphatic structure will illustrate the principle by examining how the former is rendered from English into two other genealogically closely related Slavonic languages, Bulgarian and Russian, and the various phonological and lexico-grammatical devices each language resorts to.

The piece of illustrative material comes from F. Scott Fitzgerald's classic *The Great Gatsby* (1925), a novel set in the get-rich-quick American Roaring Twenties or The Jazz Age, as Scott Fitzgerald himself named the period. The excerpt is from Chapter One where Nick Carraway, the voice of both the novel and the novelist, is introduced to Daisy and Miss Baker by Daisy's husband, Tom Buchanan. The quotation includes some of the preceding and some of the following text in order to provide the relevant context[2]:

The younger of the two was a stranger to me. She was extended full length at her end of the divan, completely motionless and with her chin raised a little as if she were balancing something on it which was quite likely to fall. If she saw me out of the corner of her eyes she gave no hint of it – indeed I was almost surprised into murmuring an apology for having disturbed her by coming in.

> The other girl, Daisy, made an attempt to rise – she leaned slightly forward with a conscientious expression – then she laughed, an absurd, charming little laugh, and I laughed too and came forward into the room.
>
> (1) "I'm p-paralyzed with happiness."

[2] The English language version of the text follows Fitzgerald, F. Scott (1991 [1925]) *The Great Gatsby: The Authorized Text,* Scribner Paperback Fiction: New York. The Bulgarian text is from Фицджералд, Ф. Скот (1966) *Великият Гетсби,* Народна култура: София. Translated from English by Nelly Dospevska. The Russian text is from Фицджералд, Ф. Скот (1977) *Избранные произведения в трёх томах,* vol. 1, Художественная литература: Москва. Translated from English by E. Kalashnikovaya.

> She laughed again, as if she had something very witty, and held my hand for a moment, looking up into my face, promising that that there was no one in the world she so much wanted to see. That was a way she had. She hinted in a murmur that the surname of the balancing girl was Baker. (I've heard it said that Daisy's murmur was only to make people lean toward her, an irrelevant criticism that made it no less charming).

The Bulgarian and the Russian versions are as follows:

(2) П-парализирана съм от щастие.
P-paralyzed-fem. am from happiness

(3) На м-меня от радости столбняк нашел.
To m-me from happiness shudder found

All three versions, (1), (2) and (3), employ intonation whereby the prominent word, the nucleus – 'p-paralyzed', 'п-парализирана' (p-paralyzed-fem.) and 'м-меня' (me), respectively – is intensified through an exaggeration of the prominence-giving pitch. Yet in this case, and in general with emphasis for intensity[3], pitch cannot be regarded as the essential characteristic of intensifying expressions. Extra length in the articulation of the respective word initial syllable onsets that lengthens the respective consonants creates both according to the author and according to the translators the total – or as in the case of Russian – the partial effect of what is expected by Daisy's 'positive' exaggeration. In English, and particularly in American English since the Jazz Age, the effect is achieved by lengthening syllable onsets most often of stressed syllables. The artificiality, and hence the inadequacy of both Bulgarian and Russian renditions, comes from the literal transposition of the linguistic pattern from English. In Bulgarian and in Russian the emphatic effect is more

[3] Coleman distinguishes between two kinds of emphasis which he terms *emphasis for prominence* and *emphasis for intensity*. "The first kind of emphasis may be defined as that manner of utterance which marks any word or phrase as of greater importance than its neighbours… The other kind … may be defined as that manner of utterance which imparts an added degree of intensity to some part of the idea represented by a word." (Coleman 1914: 11). The ultimate effect is, as Ščerba (1963: 132) puts it, conveying an immediate emotion (непосредственное чувство) rather than a manifestation of the communicative intention of the speaker.

frequently conveyed by lengthening the vowel peaks since neither language utilizes functionally the phonological tense/ lax (long/ short) opposition, and the most natural equivalence both languages can resort to is vowel lengthening (cf. Matusevič 1976; Tilkov 1981). The semantic effect of lengthening is either *prosodic meiosis* or *prosodic hyperbole*. Thus, the adequate functional equivalent in Bulgarian to the English (1) should read:

(4) Парализи-ирана съм от щастие.

The Russian version (3), apart from consonant lengthening resorts to a lexical device and uses a phraseological unit, i.e. a structurally separable language unit with a completely or partially transferred meaning. The phraseological unit itself is emphatic, and in combination with prosodic lengthening and a higher F0 excursion, the effect amounts to unnecessarily affected tautology. The economy principle should be observed even with such iconic phenomena as prosodies.

Broadly speaking, if the figure of speech *synecdoche* is taken in its original, etymological meaning, namely 'accepting a part as responsible for the whole (*'pars pro toto'*), or vice versa (*'totum pro parte'*), then (1), (2) and (3) can be taken as exponents of synecdochic semantic relations operating on sentence level.

4. Intonation and grammar: intonation as a case marker

In his monograph *Intonation and the Grammar in British English* (1967) Halliday axiomatically affirms that ""intonational" systems operate at many different places in the grammar" (Halliday 1967: 10–11). Furthermore, "it is the requirements of the grammar that set the limits of delicacy on the phonological statement" (Halliday 1967: 11). If the fall-rise dichotomy, as Cruttenden (1981) claims, is taken as a universal but is at a higher level of abstraction, then the manifestations of their meanings at lower level can be of lexical, grammatical, discoursal or attitudinal character. Phylogenetically and ontogenetically grammar takes the primacy and if meaningful instances of manifestations of intonation are to be searched for it is there, in grammar, especially in syntax, that intonational tones are playing their primitive dance with one foot as Bolinger (1964: 844) once metaphorically put it. Thus, to answer the basic methodological question of whether a tone is *meaningful,* is to verify, as Halliday claims whether "it is exploited *somewhere* in the grammar or

lexis of the language" (Halliday 1967: 12). This is because "there is no difference *in the way they work in the grammar* between systems with direct phonological exponence, such as those carried by intonation, and those expounded indirectly through a long chain of grammatical abstraction" (Halliday 1967: 10).

In a previous paper (cf. Phillipov 2011) I came up with the hypothetic claim that intonation seems to be 'wandering' around the category of case as a meaningful sign. In that way it gets rid of what was metaphorically termed *the Cheshire cat's syndrome*. The example that was presented was from German:

(5) Herr Müller schickte dir das Buch und nicht Anna.

If the nucleus in (5) is associated with *Müller*, then it is contrasted with *Anna,* the former being a subject, and the English equivalent will be 'Mr Müller sent you that book, not Anna'. If the nucleus is associated with *dir* ('to you'), then it is contrasted with *Anna* as indirect object, and then the translation equivalent will be 'Mr Müller sent a book to you, not to Anna'. In these two examples, it was claimed, the only means of resolving case relation is intonation, since the proper name *Anna* cannot be (pre)modified by an article, which in German is one of the exponents of the case system of nouns. If *Anna* in the latter instance is an exponent of a dative case form, then it is only because in the grammatical system of German there is dative as the unmarked exponent of indirect objects of this type. The competing mechanism is invariably at work in language. If Hjelmslev's thesis that, '[l]e cas est une catégorie, qui exprime une relation entre deux objets' Hjelmslev, 1935: 96) is accepted, intonation can be viewed as a marked exponent of case. It is not a mere competing mechanism', for it competes with a zero marker in this case; rather it is an instance of what the Russian linguist A. M. Peshkovsky referred to as the "compensatory law" of intonation:

> Чем яснее выражено какое-либо синтаксическое значение чисто грамматическими средствами, тем слабее может быть его интонационное выражение (вплоть до полного исчезновения), и наоборот, чем сильнее интонационное

> выражение, тем слабее может быть грамматическое (тоже до полного исчезновения). (Peshkovsky 1959: 181)[4]

(5) is not a mere *hapax* in the grammatical system of German, neither is a German one of the rare languages that a grammatical phenomenon of this kind can occur in. (6) presents a similar case:

> (6) Ich schenke das Bild der Tochter deiner Freundin.[5]
> I present the picture to/ of the daughter of/ to your friend

(6) is ambiguous. If the nucleus is associated with *der Tochter*, an indirect object, then intonation, i.e. the pitch movement is an exponent of the dative case and *deiner Freundin*, a postmodifier, and a genitive form, of the former. In this case there will also be a break between the first IP *Ich schenke das Buch* and the second IP *der Tochter deiner Freundin.* Alternatively, if *der Tochter* is a postmodifier, hence a genive form, of the direct object *das Bild*, then the nucleus in the second IP is associated with the dative form *deiner Freundin*. In this case a break will occur between the NPs *der Tochter* and *deiner Freundin.* (6) exemplifies syncretism of case forms, disambiguation being realized by intonation systems: in this case tonicity and tonality.

Intonation interacting so intimately with case seems to be a typological characteristic of synthetic languages. Amongst the first to tackle indirectly the issue was the Roman rhetor Quintilian. In his *Institutio oratoria.* In Book VII, Chapter 9, which deals with types of ambiguity, Quintilian provides evidence from Latin exemplifying the variety of ways the systems of intonation can interact with syntax. (7) is presented as a felicitous example illustrating the point at issue:

> (7) Aio te, Aeacida, Romanos vincere posse.

(7) is ambiguous: it can either mean 'I assert that you, son of Aeacis, can defeat the Romans' or 'I assert, you son of Aeacis, can be defeated by the Romans'. It can be conjectured that in Latin the disambiguation depended

[4] The clearer a syntactic meaning is realized with purely grammatical means, the weaker will be its intonational expression (till its complete disappearance), and vice versa, the stronger the intonational manifestation, the weaker will be the grammatical one (till its complete disappearance). [My translation –VP].
[5] The example is from Admoni (1955: 94).

on whether the nucleus was on *te* or on *Romanos.* In the former version *te*, a simple accusative form, is the subject of the subordinate clause *Romanos vincere posse.* The latter version exemplifies the structure *accusativus cum infinitivo,* which in this instance is a manifestation of *accusativus duplex* (*te* and *Romanos*), whereby *Romanos* is the deep subject of the clause although morphologically it is in the accusative.

Russian, a highly inflected Slavonic language, can serve as an ample source of evidence. Thus note the ambiguities in the following examples, the latter can be disambiguated only by activating the systems of intonation: tonality, tonicity, and tone:

(8a) Оля,/ чувствовала мать,/ не зря вызвала ее.
(Olya (NOM),/ felt the mother (NOM)/ did not call her in vain)
(8b) Оля чувствовала:/ мать незря вызвала ее.
(Olya (NOM) felt: / the mother (NOM) did not call her in vain)

For the disambiguation of (8a) and (8b) the systems of tonality, tonicity and tone are at work, with tonality taking the lead. Although in both (8a) and (8b) both *Olya* and *mother* are on the surface representation marked for Nominative, syntactically (8a) is an example of a subject of a parenthetical clause, whereas (8b) is the subject of an object clause

One more example from Russian will suffice to illustrate the complexity of the phenomenon:

(9a) Надо учиться,/ работать/ и отдыхать.
(One has to study, / work/ and relax)
(9b) Надо учиться работать и отдыхать.
(One has to learn how to work and relax)

(9a) is a compositional syntactic chain of three independent infinitival structures – ‘учиться’ (to study), ‘работать’ (to work), and ‘отдыхать’ (to relax) and is also a realization of three IPs, each one mapping with a syntactic structure. (9b) is one IP whereby the nucleus is associated with the first infinitival form ‘учиться’. This is a case of *infinitivus obiecti directi* with transitive verbs whose meaning goes to the meaning of the respective abstract noun (‘work’, ‘rest’, etc.). Indeed, on a deeper level of representation two independent VPs are to be posited: *учиться работать* and *учиться отдыхать.* This fact is substantiated by diachronic evidence: in the Indo-European languages all infinitival forms originate

from oblique cases of *nomina actionis* or of *verba abstracta.* There existed a productive tendency in those historically remote days that the so-called *nomina actionis* should get structurally closer to the verbs, thus giving rise to the category of the infinitive. Such piece of evidence shows that it is worth searching for a semantic/ prosodic interface between case and intonation.

Views in linguistics on the category of case seem to be poles apart. At one extreme is Jespersen's position that "cases form one of the most irrational part of language in general" (Jespersen 1924: 186). Jespersen finds theoretical support in H. Paul and quotes Paul's famous dictum: "Die Kasus sind nur Ausdrücks Mittel,…" (Jespersen 1924: 186, f.n. 1)[6] and the category is absent in the so-called analytical languages. At the opposite extreme is L. Hjelmslev (1935). At present, the tendency seems to support Hjelmslev's standpoint. In fact, over the course of the last forty years the generative paradigm has accumulated ample evidence in support of such a view. The influence is palpably felt not only in Chomsky's own GB theory, but also in the Case Grammar of Fillmore, the Relational Grammar of Perlmutter and Postal, and the Lexical Functional Grammar of Bresnan. A common theme that unites these approaches is that there is a limited set of universal semantic roles such as agent, experience, patient and instrument which are manifested by deep semantic cases. Broadly speaking, case is understood to mean "a system of marking dependent nouns for the type of relationship they bear to their head" (Blake 1994: 13). And all means of grammatical relations can be forms of case marking. This seems to be the case because "the category of case was … the most peculiarly grammatical of all the traditional categories of inflexion, for it had no counterpart in the sister sciences of logic, epistemology and metaphysics" (Lyons 1968: 289).

When selecting a case grammar model to account for intonation the following two criteria should be taken into consideration:

(a) the case grammar should provide the optimal explanatory model;

(b) this should be done in the most economical way.

After a careful scrutiny of the various versions of Case Grammars, the model that was ultimately adopted was that of Relational Grammar

[6] The cases are only means of expression.

originally developed by Perlmutter and Postal in the early 1970s. An obvious advantage of the theory is the neutral starting point that the grammatical relations are viewed as undefined primitives. The first group of primitives consists of *terms* which are semantically heterogeneous and are explicitly related to the grammatical relations subject, direct object and indirect object. The other group is formed by the semantic roles of the so-called *obliques* and includes the locative, the benefactive and the instrumental. The first group includes purely syntactic relations, whereas the second are semantic. The grammatical relations are hierarchically arranged in the following way:

(10)	subject	direct object	indirect object	obliques
	1	2	3	

In terms of a broad level model, the theory is multistratal, i.e. the dependents of the verb may display different relations in different strata. At the initial strata, the relations are semantically determined. Thus, (5), here repeated first as (11a) in its version 'Mr Müller (NOM) sent you that book, not Anna (NOM)' reflects initial stratum relations directly:

(11a) Herr Müller	schickte	dir	das Buch	und nicht	Anna.
1		3	2		1

In its version 'Mr Müller sent a book to you, not to Anna' (5), here repeated as (11b) has the following multistratal representation:

(11b) Herr Müller	schickte	dir	das Buch	und nicht	Anna.	
1		3	2		1	initial stratum
1		2	chômeur		3	final stratum

In (11b) *Anna* can be a subject but in this case it is advanced to an indirect object and is given 3 on the representational scale. The direct object *dir* because it is a recipient and in a double-object construction it advances to 2. The evidential support which Relational Grammar adduces for taking the recipient object to be the direct object is that it can be advanced to subject by the passive. *Das Buch* is demoted to what the theory labels as chômeur. The French word *chômeur* means unemployed person and in Relational Grammar is an extended metaphor for a dependent displaced from term status by revaluation, usually because of the advancement of another dependent (in the case of (11b), it is *Anna*).

The hypothesis is that intonation will interact with syntax in the domain of the terms. In the case of the *obliques*, it seems to be more

closely linked with information structure and leave the grammatical domain altogether. In (11a) *Herr Müller r*and *Anna* are both parallelly given 1, hence the same tone on both (a fall). In (11b), however, it is *dir* and *Anna* which form a syntactically parallel structure, hence they are related in terms of gradience from 2 to 3 where in linear modification "gradation of position creates gradation of meaning when there are no interfering factors" as Bolinger once aptly put it (Bolinger 1952: 1125).

5. Conclusion

With the advent of generative grammar, more definitive light was shed on many controversial issues in linguistics. Intonation, the 'Cinderella' of the discipline is still queuing, and while doing so goes up and down somewhat erratically the levels of academic linguistic descriptions. This paper has made an attempt at accommodating the 'wandering Jew' of linguistics to its original Promised Land, and anchor it to its house – the grammatical category of case.

References

Admoni, V.G. 1955. *Vvedenie v sisntaksis sovremennogo nemetskogo yazyka.* Moscow: Izdatel'stvo literaturyna innostrannyh yazykah.

Benveniste, E. 1966. *Problèmes de linguistique générale.* Paris: Gallimard.

Blake, B. J. 1994. *Case.* Cambridge: CUP.

Bolinger, D. L. 1952. Linear Modification. *Publications of the Modern Linguistic Association of America* 67: 1117–1144.

Bolinger, D. L. 1964. Intonation as a Universal. In *Proceedings of the Ninth Congress of Linguists, Cambridge, Mass. 1962,* edited by H.G. Lunt, 833–848. The Hague: Mouton.

Brazil, D. 1981. The Placement of Intonation in a Discourse Model. In *Studies in Discourse Analysis,* edited by M. Coulthard and M. Montgomery, 146–157. London: Routledge and Kegan Paul.

Coleman, H. O. 1914. Intonation and Emphasis. *Miscellanea Phonetica* I: 6–26.

Coulthard, M. & M. Montgomery (eds.). 1981. *Studies in Discourse Analysis.* London: Routledge and Kegan Paul.

Cruttenden, A. 1981. Falls and Rises: Meanings and Universals. *Journal of Linguistics* 17: 77–91.

Gramatika na suvremenniya bulgarski knizhoven ezik, Tom I: *Fonetika.* 1982. Sofia: BAN.
Gramatika na suvremenniya bulgarski knizhoven ezik, Tom III: *Sintaksis.* 1983. Sofia: BAN.
Halliday, M. A. K. 1961. Categories of the Theory of Grammar. *Word* 17: 241–292.
—. 1967. *Intonation and Grammar in British English.* The Hague: Mouton.
Halliday, M. A. K. & W. S. Greaves. 2008. *Intonation in the Grammar of English.* London and Oakville: Equinox.
Hjelmslev, L. 1935. *La catégorie des cas: Etude de grammaire générale I.* Copenhagen: Munksgaard.
Jespersen, O. 1924. *The Philosophy of Grammar.* London: George Allen and Unwin.
Ladd, D. R. 2008. *Intonational Phonology.* Second edition. Cambridge: CUP.
Lyons, J. 1968. *Introduction to Theoretical Linguistics.* London and New York: CUP.
Martinet, A. 1964. *Elements of General Linguistics.* London: Faber and Faber.
Matusevič, M. I. 1976. *Sovremennyi russkiy yazyk: Fonetika.* Moscow: Vysshaya shkola.
Molhova, J. 1981. The Phenomenon Fluctuation in English. *Philologia* 8–9: 47–56.
Pencheva, M. 1983. Za haraktera i myastoto na slovoobrazuvaneto v sistemata na ezika. *Contrastive Linguistics* 4: 54–61.
Peshkovsky, A. M. 1959. Intonatsiya I grammatika. In *Izbrannye trudy* by A. M. Peshkovsky. Moscow: Nauka.
Phillipov, V. 2012. The Sign Character of Intonation. In *Exploring English Phonetics,* edited by T. Paunović and B. Čubrović, 99–109. Newcastle upon Tyne: Cambridge Scholars Publishing.
Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. 1985. *A Comprehensive Grammar of the English Language.*London: Longman.
Ronat, M. 1984. Logical Form and Prosodic Islands. In *Intonation, Accent and Rhythm: Studies in Discourse Phonology,* edited by D. Gibbon and H. Richter, 311–326. Berlin/New York: Walter de Gruyter.
Sčerba, L. V. 1963. *Fonetika frantsuzkogo yazyka.* Moscow: Vysshaya shkola.
Selkirk, E. O. 1986. *Phonology ans Syntax: The Relation between Sound and Structure.* Cambridge, Mass./London: The MIT Press.

Tilkov, D. 1981. *Intonatsiyata na bulgarkiya ezik.* Sofia: Narodna prosveta.

INTONATION PATTERNS AND PHONETIC STEREOTYPES: NEW LIFE FOR OLD TERMINOLOGY

YULIA NENASHEVA

Outline

This paper examines approaches to prosodic research, and presents some of the results of intonation study. To examine intonation of an utterance, a study of its prosodic components is undertaken. Durational, dynamic, and tonal qualities were measured and statistically analyzed. It is found that prosodic elements comprise a complex structure of interrelated units. The arrangement of the units is predictable and carries specific meanings. Research shows that these prosodic complexes possess certain distinctive features, the set of which identifies them as intonation patterns that serve as models in speech production. Through these intonation patterns, phonetic stereotypes are actualized in speech. The meaning of the utterance is expressed through the arrangement and interaction of prosodic elements in an intonation pattern.

1. Introduction

The way prosodic units act and interact to represent and express mental concepts through meanings of speech has always been viewed as one of the most problematic research areas in the study of intonation. In recent years there has been an increasing body of research on the subject of prosodic features of various types of English and other languages (Grabe 2001, 2002; Gussenhoven 2002; Grabe and Karpinski 2003; Braun et al. 2006; Prieto et al. 2012). The abundance of information on prosodic elements has contributed a lot to the phonetic knowledge, but to understand how linguistic units serve as means, through which mental

images and ideas are conveyed, interdisciplinary research and further systematization of its findings should be undertaken.

This paper presents preliminary results of experimental phonetic research of "basic" intonation patterns. The results of the study are considered in terms of describing possible intonation patterns used to represent basic communicative meanings of utterances.

2. Stereotype in linguistics

One of the long-known concepts used by linguists is that of a stereotype. It is being studied in cognitive linguistics nowadays, with reference to how stereotyping influences speech production mechanisms. It seems necessary to look upon findings of cognitive linguistics from the traditional point of view. Integrating traditional terminology into new approaches to linguistics will help to categorize linguistic units investigated by cognitive linguistics and fill old terminology with new meaning.

Definition given by Crystal describes the stereotype as "a set of properties regarded by a community of speakers as characterizing typical members of a category, or linguistic variable which is a widely recognized characterization of the speech of a particular group" (Crystal 2008:452). This understanding of the stereotype is close to H. Putnam's theory, which presents stereotypes as the basic semantic knowledge, the minimal amount of semantic knowledge that a language user is supposed to have if he considers himself a member of the linguistic community (Putnam 1975).

There is one connotation frequently present in the meaning of the term "stereotype", which may confuse linguists: the term is used to denote "a set of (often pejorative) characteristics". Different views on stereotypes make it difficult to give a single definition of the concept. Many definitions speak of the stereotype as a unit that has little or no productivity and represents caricature or outdated linguistic features of the speech community. Definition of the stereotype given by Crystal points out that the term may not reflect accurately the speech of the group it is supposed to represent, or it allows for inaccurate beliefs on the part of the speech community. Nonetheless, the definition stresses the point that knowledge of the stereotype is required for semantic competence in the language (Crystal 2008).

A stereotype is a unit which helps people to systematize and organize their knowledge of the world and use this experience to predict future experiences on the basis of their similarity. This understanding of

stereotype is close to Lakoff`s definition of cluster models. And as Lakoff says "the category can be structured by a cluster model consisting of a number of different subcategories, one of which comes to stand for the category as a whole and serve as a cognitive reference point, setting up norms and expectations against which other members of the category are evaluated and assessed", these categories are complex structured systems of knowledge: language prompts for the construction of mental spaces in ongoing discourse (Lakoff 1987: 74). Thus, linguistic stereotype can be looked upon as the most common, most central meaning, the unit, which serves as a helping tool in discourse (Maas & Arcuri 1996).

In the view of Ladd, the more general idea of "stereotype" or "routine" has been put forward, and very convincingly (Ladd 1980). Ladd introduces the term "stylized intonation", and this system of intonation is not restricted only to stylized or stereotyped forms of speech but is also found in reading pronunciation and conversational speech. Ladd also notes that the core meaning of stylized intonation is "predictability". Similar idea is presented by sociophonetics: a new approach to phonology with a strong phonetic bent. Its main hypothesis is that the phonological knowledge of a speaker is constructed upon a stock of phrases that they have heard other people say (Pierrehumbert 2003; Coulson 2006). This has been proved correct by child speech and language acquisition research.

Linguists have known for a long time that by the time children reach the age of five or six, they are proficient speakers of their native language. On the other hand, when a child or an adult learns a second language, they usually encounter some kind of problem: second language listeners feel uneasy about using stereotypical knowledge from their own culture while interpreting what is said in another language. They also fail to notice when they have not properly understood what someone says, or blame themselves for not having understood native speakers. Pierrehumbert notes that acquisition of phonological categories is gradual, the gradualness of acquisition is correlated with the frequency of the variants and with what groups within the community use more often (Pierrehumbert 2003: 185; see also Foulkes & Docherty 2006).

The idea that it is possible to possess a concept while not knowing, or being mistaken about its properties through interference of a different language in second-language learners, represents one of the most important problems in teaching phonetics, especially intonation.

In communication there are always two participants, and the relationship between them defines the speech strategy and the choice of linguistic means. The speaker makes judgments about how well the

addressee will understand the meaning transported by the speech. When listening to a speaker, listeners often make stereotypical judgments about the speaker, not only on the basis of the speaker's self-presentation, but also on the basis of perceived speech.

Gussenhoven identifies three biological codes, conditioned by human physiology. These codes represent natural form-function relations, and they are, to a different degree, employed by speech communities. Such exploitation of these phenomena will to some extent be conventionalised within speech communities. Speakers using the codes do not have to create specific physiological conditions. It is enough to create the effects. These effects are controlled during phonetic implementation (Gussenhoven 2002).

Alan Bell quotes Bakhtin: "…the utterance of the person to whom I am responding …is already at hand, but his response (or responsive understanding) is still forthcoming. When constructing my utterance, I try actively to determine this response. Moreover, I try to act in accordance with the response I anticipate, so this anticipated response, in turn, exerts an active influence on my utterance… When speaking I always take into account the apperceptive background of the addressee's perception of my speech... These considerations also determine my choice of a genre for my utterance, my choice of compositional devices, and, finally, my choice of language vehicles, that is, the style of my utterance" (Bell 2007:107).

According to another linguist, utterances predict and arrange social communication occasion in which they occur; they predict social relations as effects of their occurrence (Agha 2007). This aspect of speech involves knowing the social and cultural rules for using a language (Maas and Arcuri 1996). It depends on how many social features the speaker and listener share (Clark 1998). Hymes developed a list of sociolinguistic awareness elements that are involved when speakers communicate in particular speaking situations. They may be as follows: type of event, topic, purpose/function, setting, key/emotional tone, participants, message form/content, act sequence (the ordering of communicative phenomena), rules of interaction and norms of interpretation (cultural expectations about how talk should proceed, and what its significance is) (Hymes 1996). Thus, the speakers' perception, evaluation and interpretation processes are culture-bound, which influences their manner of pronunciation, word or prosody choice.

This brings us very close to the stereotype which on the phonological level of the language structure exists as a phonetic/phonological stereotype – linguistic awareness that is involved when speakers communicate in

particular speaking situations or express particular meanings using particular prosodic elements, or arranging them in specific order. Thus, linguistic stereotype is characterized by the following features:

- It represents basic semantic knowledge.
- It serves as a cognitive reference point and it sets up norms and expectations against which other units are evaluated and assessed.
- It is social.
- It is predictable, because it is conditioned by speaking situations.
- It is actualized in speech through specific set or arrangement of units

3. Present study

Language describes reality. Each language has means to convey messages about people, things, concepts and the relationships between them. Messages about reality are structured so that they can be processed by the parties taking part in communication. There is always correspondence between the participants of events, types of interaction they enter, and linguistic units they use. When a human being perceives speech, he recognizes intonation patterns first, and then uses their meaning to decode perceived speech. When a human being produces speech, he uses phonetic stereotypes to translate his thoughts into linguistic concepts and intonation patterns to put them into actual speech.

Speech is not as accidental and individual as it sounds. Every speaker possesses a set of social rules and expectations, which help him to perceive and decode speech of other people, even if there was no previous experience of communication between them. These social expectations and rules are structured in the speakers' phonological knowledge in clusters of phonological categories – phonetic stereotypes, which, in turn, are represented in the language through specific forms and instances of language units – intonation patterns. The existence of an intonation patterns as a bigger phonological suprasegmental unit, containing smaller prosodic units or complexes, has been mentioned by many linguists. Hirst speaks of “a remarkable consensus concerning the existence of prosodic constituents equivalent to” what he calls “Intonation Units” (Hirst 1998: 59, 71).

For an intonation pattern (as a representative of a stereotype) to be a phonological unit means to be a point of reference against which all other prosodic arrangements will be marked as “having a different

communicative meaning". Here a concept of a "basic intonation pattern" comes into existence. Such a pattern should be characterized by specific characteristics, which can be identified through systematic linguistic analysis and interdisciplinary research. The necessity of such a concept has been stated by many linguists (Hirst 1998; Kristiansen 2003). Coulson (2006: 19) terms it "default". From the phonological point of view, such an intonation pattern ("basic" or "neutral" intonation pattern) represents prosodic minimum. "Basic" intonation patterns must be semantically simplified, as the influence of additional extralinguistic factors causes the increase in the number of prosodic features in the intonation pattern of the utterance. Semantic structure of such patterns should be limited by three basic elements: reference to the object, communicative type of the utterance and "neutrality".

Crystal and Davy note that the average length of Intonation Units was five words and that 80% of the Units were less than eight words long. When utterances are longer than this, they are usually broken up into two or more Intonation Units. Utterances, which contain more than three or four pitch accents in a single Intonation Unit, are quite rare in spontaneous speech (Crystal & Davy 1969).

It is necessary to explain the term "neutrality". The dictionary explains "neutral", as "of no particular kind, characteristics". So, everything that deviates from usual is identified as "specific". However, "neutrality" of the utterance is its structural property, represented on all levels of linguistic system (Halliday 1967). Neutral utterances are minimally dependent on the context of the situation. They are characterized by standard grammatical structure and absence of any specific prominence of any element. These utterances are not involved in delivering any emotional or emphatic meaning. "Neutral" utterances serve as "weak" members of opposition. Upon their background, the speakers identify and distinguish variants, which express specific communicative meanings (Kristiansen 2003, 2010; Coulson 2006).

Study and description of such "basic", "neutral", "default" intonation patterns will help to understand how mind and language work, identify means through which mental images and ideas are conveyed, classify linguistic models and meanings they communicate.

3.1. Results

The research involves analysis of 486 utterances presented by the project IViE (Grabe *et al.* 2002) at Oxford University Phonetics

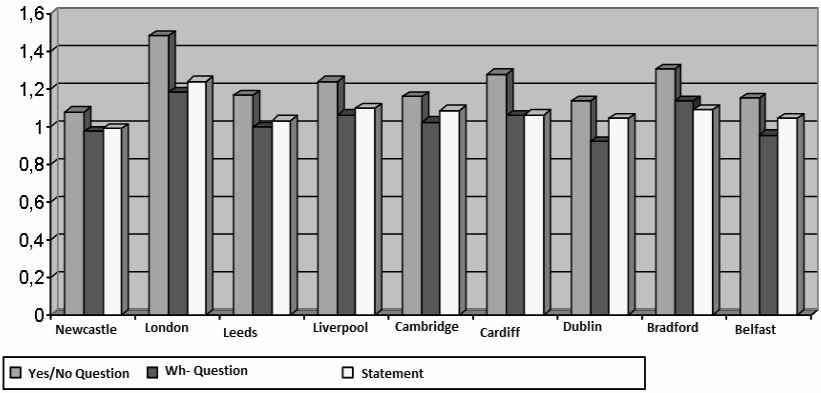
Laboratory. With the help of computational analysis (PRAAT 5.2.03© 1992–2010 by P. Boersma and D. Weenik), inventory, distribution and dynamics of prosodic parameters in the intonation pattern was investigated. The data acquired through analysis has been evaluated with the help of statistics (SigmaStat ©).

Number of prosodic parameters (durational, dynamic and pitch) were analyzed and the results were arranged into tables. The data was assessed through statistic analysis and the results were presented in diagrams. The results of the analysis were evaluated against physiological limits of perception (Stevens 2000; Johnson 2003).

Preliminary results show that prosodic complex – intonation pattern – of a "neutral" utterance is characterized by the following features:

Such an utterance may contain from five to seven syllables. The duration of a "neutral" utterance is usually no longer than 1–2 seconds as shown in Figure 1.

Figure 1. Diagram of duration of "neutral" utterances



Statistic evaluation shows that highest degree of variation in duration of the utterance is found within *Yes/No questions*. *Wh-questions* and *Statements* tend to have similar duration.

The length of its functional parts is distributed in the proportion 60%:40%, where the prenuclear part is longer than both Nucleus and Tail (Figures 2–4). The Pre-Head, which contains unstressed syllables, is shorter than any other part of the intonation pattern.

According to Figures 2–4, Head in all utterances appears to be the longest part of the Intonation Unit as it contains more syllables. Duration of Tail of the Intonation Units in all utterances does not reveal any statistically significant differences.

Proportional sum total duration of Pre-Head and Head in *Statements* does not reveal any significant differences with the proportional duration of Head in Wh-questions as shown in Figures 3–4. Proportional duration of Tail in all utterances does not reveal statistically significant differences.

Figure 2. Proportional duration of functional parts of the Intonation Unit (Yes/No Question)

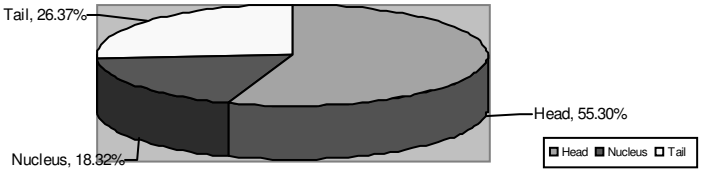


Figure 3. Proportional duration of functional parts of the Intonation Unit (Wh-Question)

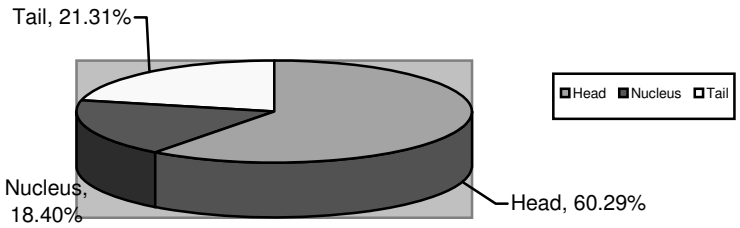
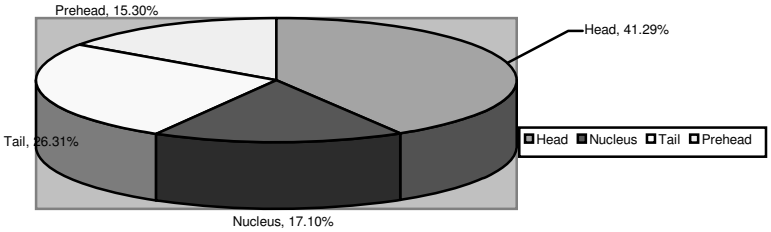


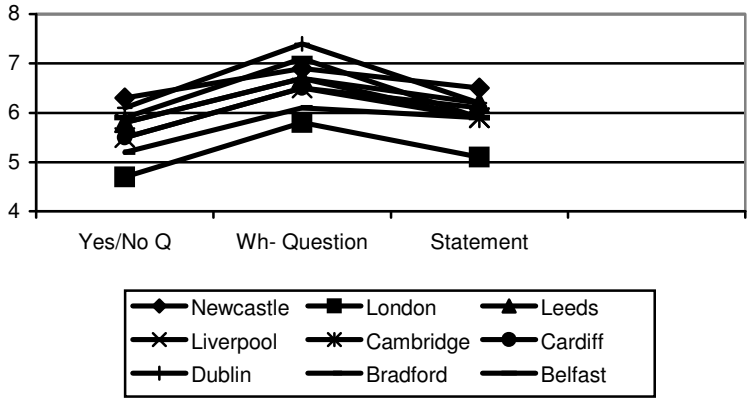
Figure 4. Proportional duration of functional parts of the Intonation Unit (Statement)



In all utterances proportional duration of Nucleus does not reveal any statistically significant differences.

Utterances used in research are characterized by the normal tempo (4–7 syllables per minute). *Wh-questions* are characterized by higher tempo than other utterances. Within regional groups of speakers statistically significant differences were found (Figure 5).

Figure 5. Line plot of tempo characteristics of "neutral" utterances.



Average syllable duration parameter was used to establish the degree of syllable compression within the utterance. The results confirmed that parts of the utterance considered less relevant to the goal of communication (prenuclear), is pronounced slightly quicker, though the tempo remains within the normal range.

The research also confirmed that average syllable duration parameter increases towards the end of the utterance (Figures 6–8). These findings fall in line with those about the behaviour of tempo characteristics, which indicates that tempo in all utterances decreases towards the end of the utterance

Figure 6. Line plot of average syllable duration within functional parts of "neutral" utterances (Yes/No Question)

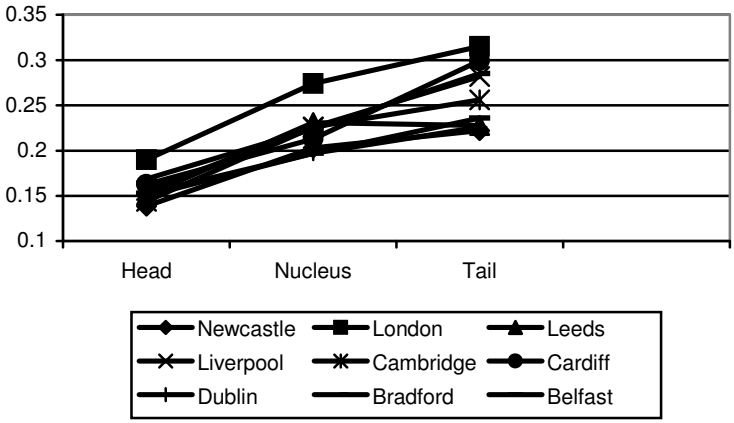


Figure 7. Line plot of average syllable duration within functional parts of "neutral" utterances (Wh-Question)

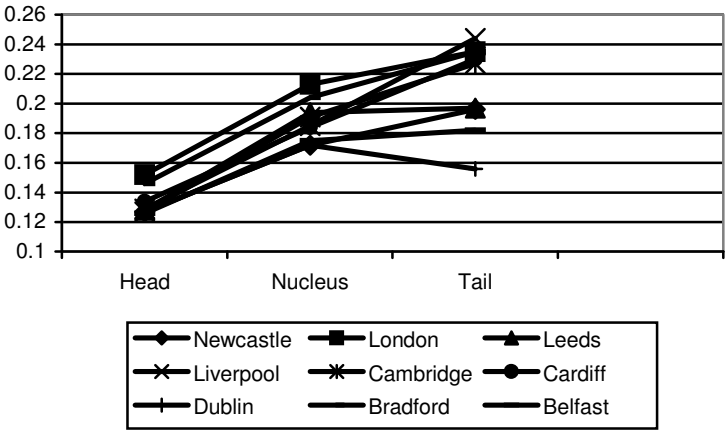
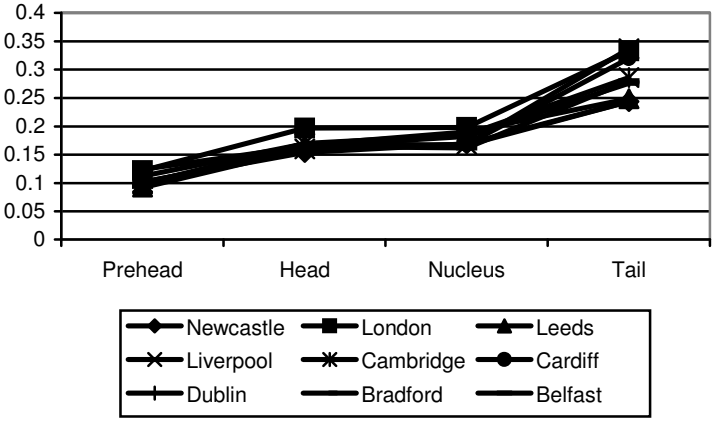


Figure 8. Line plot of average syllable duration within functional parts of "neutral" utterances (Statement)



As shown in Figures 6–8, Heads in *Yes/No questions* and *Wh-questions* are pronounced faster than other functional parts of the utterance. Average syllable duration in the Nucleus does not reveal any significant differences in all types of utterances.

"Neutral" utterances are characterized by a peculiar distribution of prosodic parameters within the utterance: parts with greater acoustic parameters alternate with those with lesser acoustic parameters. It can be seen in Figures 6–8. And in all utterances tempo decreases towards the end of the utterance.

Dynamic and pitch parameters were measured and evaluated within the functional parts of the utterances. Dynamic parameters of utterances used in the present research follow the same pattern. Their distribution has a regular, rhythmic character: sections with greater acoustic parameters alternate with sections with lesser acoustic parameters. In all utterances intensity decreases towards the end of the utterance (Figures 9–14).

Figure 9. Line plot of intensity within functional parts of "neutral" utterances (Yes/No Question)

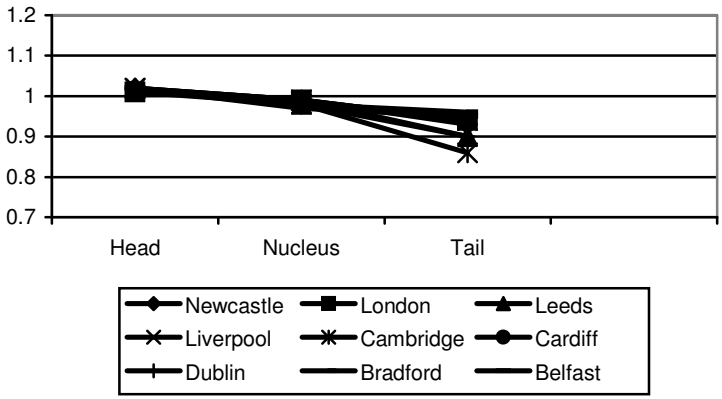


Figure 10. Line plot of intensity within functional parts of "neutral" utterances (Wh-Question)

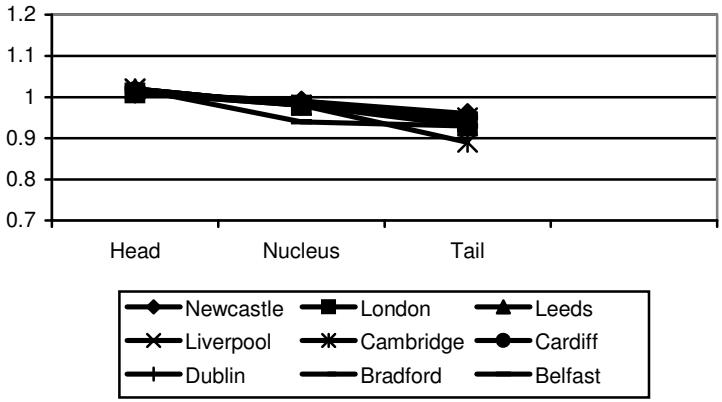
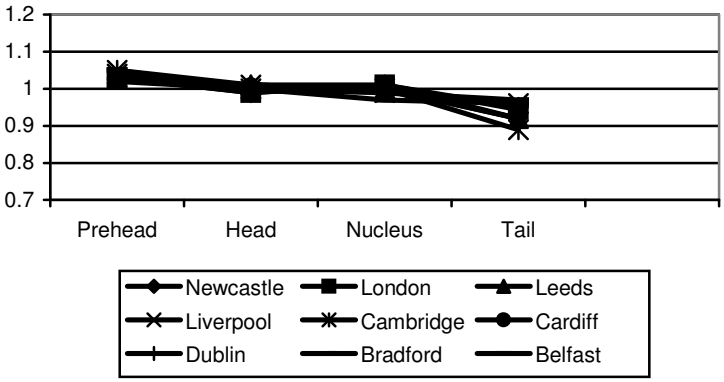


Figure 11. Line plot of intensity within functional parts of "neutral" utterances (Statement)



Pitch parameters of utterances used in the present research follow the same pattern (Figures 12–14).

Figure 12. Line plot of pitch characteristics within functional parts of "neutral" utterances (Yes/No Question)

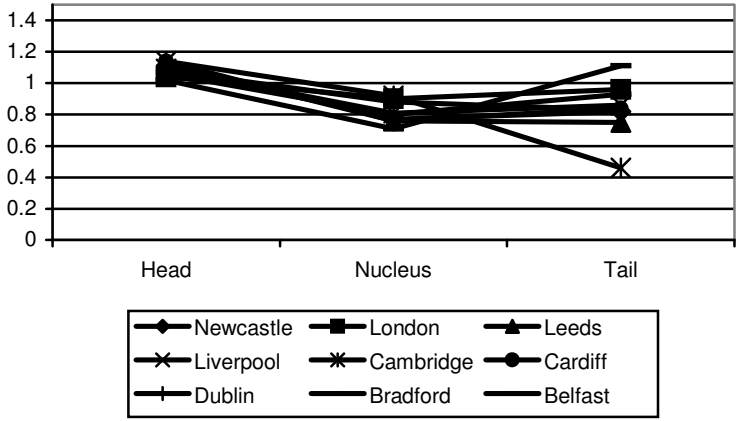


Figure 13. Line plot of pitch characteristics within functional parts of "neutral" utterances (Wh-Question)

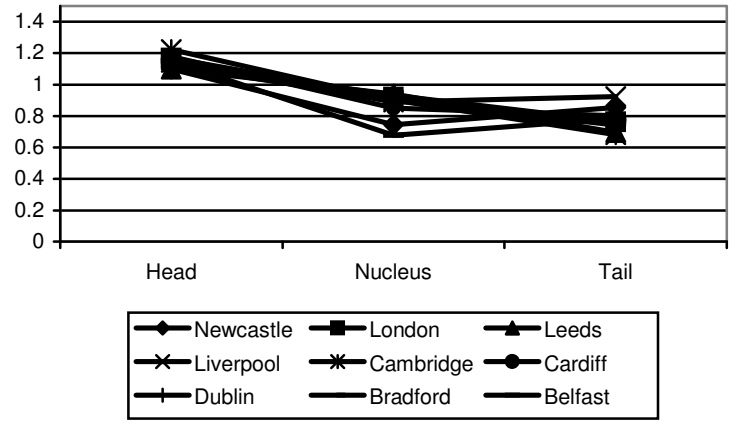
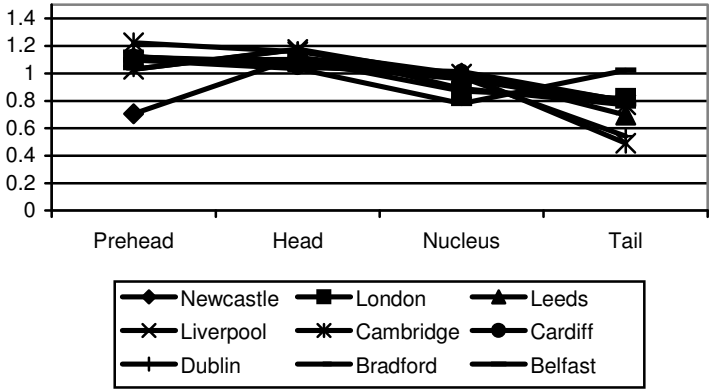


Figure 14. Line plot of pitch characteristics within functional parts of "neutral" utterances (Statement)



Their distribution has a regular, rhythmic character: sections with greater acoustic parameters alternate with sections with lesser acoustic parameters. In all utterances pitch decreases towards the end of the utterance. There are no peaks of loudness or tone prominence other than the final tonic – the final stressed syllable.

4. Discussion and conclusion

What does this research tell us about phonetic stereotypes and intonation patterns of "neutral" utterances? Statistically significant differences were found within groups of data acquired in the analysis, which is not a surprise. Differences between regional varieties of English were described by many researches (Wells 1982; Chambers and Trudgill 1998).

Within present research, the same model is found in all regional varieties involved in the study. It throws a different light on the concept of an "intonation pattern". It shows that intonation pattern is not sum total of separate prosodic elements, order, in which these elements occur and interact, is of utmost importance. The utterances are characterized by the same manner of distribution of prosodic parameters. In all utterances all prosodic parameters – durational, dynamic, and pitch – decrease to the end

of the utterance. Such utterances are distinguished by a smooth movement of prosodic elements and corresponding acoustic parameters.

All prosodic parameters alternate occurrences of lesser and greater degree, which produces specific rhythmic structure of the utterance. Such specific structure of utterances proves that there is a phonetic stereotype of a "neutral" utterance. And to all speakers such stereotype is realized through specific arrangement of prosodic elements – smooth, uninterrupted decrease of acoustic characteristics to the end of the utterance.

Further investigation into intonation structure of utterances of different communicative types or meanings will help to describe intonation patterns used to convey linguistic stereotypes. Such study will make it possible to understand how human mind works to understand and explain boundless events in the world around us.

Acknowledgements

I would like to thank the following people: my teacher L. I. Shvydkaya, who encouraged me, Biljana Čubrović, who was so considerate to invite me to BIMEP 2012, and my family, who patiently did all the housework while I was completing my research.

References

Agha, A. 2007. *Language and Social Relations*. Cambridge: CUP.

Bell, A. 2007. Style in Dialogue: Bakhtin and Sociolinguistic Theory. In *Sociolinguistic Variation: Theories, Methods, and Applications*, edited by R. Bailey and C. Lucas, 90–109. Cambridge: CUP.

Bolinger, D. 1986. *Intonation and its Parts: Melody in Spoken Speech.* Stanford, CA: Stanford UP.

Braun, B., G. Kochanski, E. Grabe & B. S. Rosner. 2006. Evidence for attractors in English intonation. *Journal of the Acoustical Society of America* 119 (6): 4006–4015.

Chambers, J. K. & P. Trudgill. 1998. *Dialectology.* Second edition. Cambridge: CUP.

Clark, H. 1998. Communal Lexicons. In *Context in Language Learning and Language Understanding*, edited by K. Malmkjær & J. Williams, 63–88. Cambridge: CUP.

Coulson, S. 2006. *Semantic Leaps: Frame-shifting and Conceptual Blending in Meaning Construction.* Cambridge: CUP.
Crystal, D. 2008. *A Dictionary of Linguistics and Phonetics*. Sixth edition. Oxford: Blackwell.
Crystal, D. & D. Davy. 1969. *Investigating English Style*. Minneapolis: Indiana UP.
Foulkes, P. & G. Docherty. 2006. The Social Life of Phonetics and Phonology. *Journal of Phonetics* 34: 409–438.
Geerarts, D. 2008. Prototypes, Stereotypes, and Semantic Norms. In *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*, edited by G. Kristiansen and R. Dirven, 21–45. Berlin: Walter de Gruyter.
Grabe, E. & M. Karpinski. 2003. Universal and language-specific aspects of intonation in English and Polish. *Proceedings of the 15th International Congress of Phonetic Sciences*, 3–9 August, Barcelona, Vol. 1, 1061–1064.
Grabe, E., B. Post and F. Nolan. 2001. Modelling intonational Variation in English. The IViE system. *Proceedings of Prosody 2000,* edited by S. Puppel and G. Demenko, 51–57. Poznan: Adam Mickiewitz University.
Grabe, E. & B. Post. 2002. Intonational Variation in English. *Proceedings of Speech Prosody*, edited by B. Bel and I. Marlien, 343–346. Aix-en-Provence: Laboratoire Parole et Langage, Université de Provence.
Gussenhoven, C. 2002. Intonation and interpretation: phonetics and phonology. *Proceedings of Speech Prosody*, edited by B. Bel and I. Marlien, 47–57. Aix-en-Provence: Laboratoire Parole et Langage, Université de Provence.
Halliday, M. A. K. 1967. *Intonation and Grammar in British English.* The Hague/Paris: Moulton.
Hirst, D. 1998. Intonation in British English. In *Intonation Systems: A Survey of Twenty Languages*, edited by D. Hirst and A. Di Cristo, 60–82. Cambridge: CUP.
Hymes, D. H. 1996. *Ethnography, Linguistics, Narrative Inequality: Toward An Understanding Of Voice.* London: Taylor and Francis.
Jassem, W. 1952. *Intonation of Conversational English (Educated Southern British).* Wrocław: Nakładem Wrocławskiego Towarzystwa Naukowego.
Johnson, K. 2003. *Acoustic and Auditory Phonetics*. Oxford: Blackwell Publishing.
Kristiansen, G. 2003. How to Do Things with Allophones: Linguistic Stereotypes as Cognitive Reference Points in Social Cognition. In

Cognitive Models in Language and Thought: Ideology, Metaphors and Meanings, edited by R. M. Frank and M. Putz, 69–123. Berlin: Walter de Gruyter.

Kristiansen, G. 2010. Lectal Acquisition and Linguistic Stereotype. In *Advances in Cognitive Sociolinguistics*, edited by D. Geerarts, G. Kristiansen and I. Peirsman, 225–265. Berlin: Walter de Gruyter.

Ladd, D. R. 1980. *The Structure of Intonational Meaning: Evidence from English.* Bloomington, IN: Indiana UP.

Lakoff, G. 1987. *Women, fire, and dangerous things: what categories reveal about the mind.* Chicago: The University of Chicago Press.

Maas, A. & L. Arcuri. 1996. Language and Stereotyping. In *Stereotypes and Stereotyping*, edited by C. N. Macrae, C. Stangore and M. Hewstone, 193–227. New York: Guilford Press.

O'Connor, J. D. 1998. *Better English Pronunciation.* Cambridge: CUP.

O'Connor, J. D. & G. F. Arnold. 1961. *Intonation of Colloquial English.* London: Longman.

Pierrehumbert, J. B. 2003. Probabilistic Phonology: Discrimination and Robustness. In *Probabilistic Linguistics*, edited by R. Bod, J. Hay and S. Jannedy, 177–228. Cambridge, MA: The MIT Press.

Prieto, P., M. M. Vanrell, L. Astruc, E. Payne & B. Post. 2012. Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication* 54(6): 681–702.

Putnam, H. 1975. The Meaning of "Meaning". In *Language, Mind, and Knowledge*, edited by K. Gunderson, 131–194. Minneapolis: University of Minnesota Press.

Stevens, K. N. 2000. *Acoustic Phonetics*. Cambridge, MA: The MIT Press.

Wells, J. 1982. *Accents of English*. Cambridge: CUP.

INTONATION INTERFERENCE AND ITS IMPACT ON EFFECTIVE COMMUNICATION BETWEEN NATIVE/NON-NATIVE SPEAKERS

OKSANA PERVEZENTSEVA

Outline

The purpose of the research is to advance understanding of the ways prosody affects communication between native and non-native speakers in situations of artificial bilingualism. Specifically we investigate which communicative-pragmatic types of utterance are mostly subjective to interference and thus cause misunderstanding or break of communication.

The research is based on the zone conception of intonation which can help determine the prosodemic status of the corresponding intonation patterns and to specify conditions of communication success or failure in the situation of artificial bilingualism.

The empirical data prove special sensitiveness of native speakers to the inaccurate use of intonation patterns and show that deviations from the intonation models are perceived mostly in the emotional-modal aspect. The application of a specific prosodic criterion will provide the basis for the development of L2 learners' communicative competence thus contributing to their adequate interpretation of emotional-modal connotations of utterances.

1. Introduction

Acquiring intonation peculiarities of a foreign language is one of the most complicated and least investigated problems in applied linguistics, though this problem has been given much more attention lately. Meanwhile the command of second language intonation (that is the ability to understand and produce authentic intonation patterns) seems to be very important in the process of communication as intonation is that litmus

paper that immediately reveals the person's attitude, feelings and emotions and by which people instinctively recognize their native or an "alien". Moreover, according to various data misuse of certain intonation patterns may result in misunderstanding or even break of communication. As Pike puts it "[...] we often react more violently to the intonational meanings than to the lexical ones; if a man's tone of voice belies his words, we immediately assume that the intonation more faithfully reflects his true linguistic intentions" (Pike 1945).

The reaction of native speakers to the learners' mistakes in grammar, vocabulary, and intonation is different. A native speaker admits that a foreigner can make grammar mistakes and misuse words. But intonation is perceived on a deeper instinctive level and as a consequence misuse in intonation is perceived not as a mistake but as the speaker's real attitude. Even small differences can be important. Speaking one language with the intonation pattern appropriate to another can give rise to entirely unintentional effects. For example, English with Russian intonation sounds unfriendly, rude or threatening to the native speaker of English; Russian with English intonation sounds affected or hypocritical to the native speaker of Russian. Numerous experiments carried out on the material of different languages yield similar results (see Grabe, Rosner, Garcia-Albea & Zhou 2003; Odlin 1996; Wennestrom 2001).

Most linguists agree that intonation is one of the phenomena without which one can never get rid of a foreign accent and be able to speak like a native speaker, but in teaching intonation seems to be the last aspect that is taught to students if ever at all. Due to its inherent complexity and to the difficulty in learning and mastering it, intonation was ignored in language teaching for many years. There are several reasons to explain this fact. First of all, it seems that intonation comes so naturally and instinctively that we shouldn't bother about it at all. It's a well-known fact that there are no languages spoken as a monotone, and intonation often seems to be universal. In fact, there are some aspects of intonation that are universal, but there are also certain language-specific intonation properties as well. The difference lies mostly in the intonation function, not its form, though the latter can be a challenge as well. Some linguists stress that even trained phoneticians and language teachers are at times unable to perceive intonation correctly and no wonder that most non-native users of English have the same problem with the production and perception of English intonation.

Besides all these difficulties in acquiring authentic second language intonation it turns out that a foreign accent in intonation is most stable and

hardest to get rid of compared to the other aspects of the language. That happens due to the strong interference (that is influence and transfer of the native language to the second language) of their native language. It is believed that interference is more persistent in those phenomena that are the earliest to be acquired in the native language. While the average educated non-native learner of English can attain a very high standard of grammatical accuracy in the language and master the pronunciation of its sound segments and word stress, s/he often cannot appropriately use its intonation with any reasonable degree of confidence.

The purpose of the research is to investigate which communicative-pragmatic types of utterance are mostly subjective to prosodic interference and thus cause misunderstanding or break of communication. We also try to work out principles of building a prosodic model of the structural organisation of speech acts.

2. Methodology

According to the latest trends in phonetic research on interference the prosodic systems of two contacting languages should be compared not only on the level of separate prosodic subsystems but also taking into account their systematic interaction both within an individual prosodic feature (fundamental frequency, intensity or duration) and in their combination. However, this approach suggests certain difficulties caused by the necessity to process a large amount of data. Modern mathematical-statistical methods and computer engineering help solve the problems. In intonology the use of the given methods is possible within the framework of intonometric analysis.

In general, intonometry is a special division of experimental phonetics whose task is to plan and process the results of measuring experiments by applying methods of mathematical statistics and building corresponding models of speech intonation on this basis.

It is common knowledge that intonation is a complex system consisting of a number of objects (elements) which possess certain features and interact with each other. Methods of multivariate analysis appear to be most effective in the study of this phenomenon in its complex. These methods allow to compress empirical data, aggregate them, and substitute by one, artificially created characteristic, which makes it compact and convenient for further research (Авен 1988). Taking into account the approximation of linguistic events of different levels, we use the method of mathematic modelling of the language system. The conception of a

blurred set with the central zone (nucleus) which is characterised by the clear set of features and periphery comes useful in our research.

The criterion of prosodic markedness and method of cluster analysis allow us to transform points of multidimensional space into two-dimensional space with the least possible distortion of their initial arrangement and to get a visual picture of the localisation of intonation realisations in the measured space. The zone conception of speech intonation is the basis for determining the prosodemic status of the corresponding intonation pattern, which can contribute to a brand new classification of speech acts. It is connected with the notions of cluster sample, tolerance area (the area of acceptable variability) and intonation zone (tolerance area in intonology). We view the cluster sample as a family of certain intonation realisations within the frame of a certain type of speech act, singled out according to its pragmatic meaning (Kanter, Chizhov & Guskova 1987).

According to the above-mentioned conception and depending on their location in the two-dimensional space and ability to form more or less compact clusters we single out:

– compact and isolated clusters which enclose one type of the intonation pattern; formation of the common cluster for two or more types of intonation is possible here;

– zones of dominating realisations, which unite more than half of the objects of the given group; presence of some objects belonging to the other types is possible;

– sub-clusters – two or more parts of the cluster that group objects of a certain type; which shows the existence of subtypes in the given intonation types.

We also use such criteria as the degree of cluster compactness and friability, the degree of indentation of cluster borders, presence/absence of the nucleus and cluster periphery. So, the compact location of intonation realisations of one communicative-pragmatic type of utterance proves the possibility of singling out this type of intonation in the given language and proves its prosodemic status. Intrazonality means that the given types of intonation function in speech as contextually conditioned variants of one intoneme.

Variability limits are different in different languages and break of the limits due to the bad command of the language causes communicative failures. More exact analysis of variability of intonation contours is possible on the basis of intonometric analysis, which allows to abstract

from individual peculiarities of a speaker, which is especially important in the situation of artificial bilingualism.

The proposed research focuses on the following nine basic intonation patterns: 2 general questions (one transformed from a command, the other – from a statement), special and rhetoric questions, a statement, an incomplete statement, a command, a statement-response, an exclamation. The research also provides an account of the role the gender factor plays ensuring the effectiveness of communication.

The contribution of attitude to meaning is particularly evident where lexically identical utterances have different meanings and those differences in meaning are claimed to be the result of the phonetic design of those utterances. That is the reason for selecting 3-syllable lexically identical (and quasi-identical) text phrases realised with the above mentioned nine basic intonation patterns. Moreover, a speech act is a convenient object for pragmaphonetic research, especially for mathematical-statistical processing of data, as it allows to analyse and describe its inner structure and build a model of speech communication with gradual transfer to the analysis of a full text.

The phrases were used in dialogue contexts by thirteen speakers (6 male and 7 female), first-year students before taking their phonetic course. The auditory experiment was conducted in three stages:

– self-audition,

– auditory analysis performed by experts/specialists (non-native speakers),

– auditory analysis performed by non-experts (native speakers).

In the first stage of the auditory experiment, the speakers (students) listened to their recordings and chose the most suitable (in their opinion) phrases which met the requirements of the experiment. The second stage of the experiment was conducted by the experts/specialists in English Phonetics (professors and post-graduate students of the English Department of the University). In the third stage, native speakers of the English language (non-experts) referred the given phrases to certain communicative types of utterance according to the seven-grade scale.

3. Results

The received data made it possible to make the following conclusions:

1. The conducted experiment, based on the method of cluster analysis, allows obtaining graphical displays of prosodically distorted speech thus

giving the opportunity to visually demonstrate the reasons of success and failure of speech acts in interference environment.

In the process of studying students easily cope with the realisation of the basic types of intonation, such as, for example, completed statement, general question, command and exclamation. Failures occur in the realisation of those speech acts which are semantically close to the mentioned ones (see *Figures 1* and *2*). For example, the analysis of clusterisation of statement-response and incomplete statement shows that they don't form independent intonation zones, which allows us to suppose that in interference speech such intonation types don't have their own prosodemic status. In sample realisations, these types of intonation form their own clusters both in the Russian and English languages and so have their own prosodemic status.

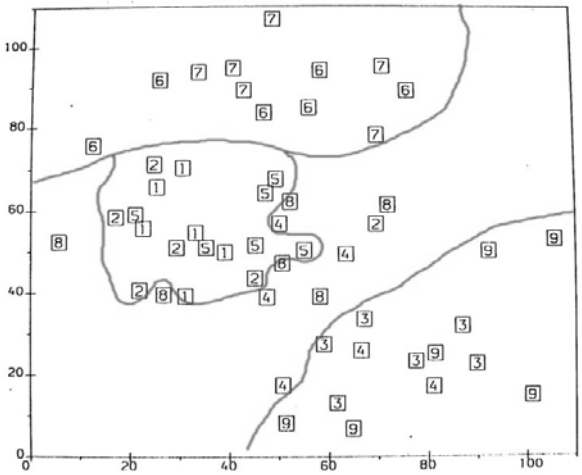
Intonation realisations of rhetoric and special questions are differentiated only in female speech and are not intonationally different in male speech. It should be noted here that rhetorical question belongs to the group of the so-called indirect speech acts, that is those speech acts in which the form and direct meaning (locution) of the phrase are not in agreement with the communicative type of the corresponding utterance (its illocutionary aim), and in which intonation plays a key role in the satisfactory conveyance by the speaker and the adequate perception by the listener of the communicative aim of the utterance. So in this case we can conclude that the realisation of indirect speech acts will fail in interference speech.

2. The research also proves that most common clusters have central-peripheral structure. The realisations of those communicative-pragmatic types of utterance which are most frequently used in communication (complete statement and special question) are located in the centre, and less important from the communicative point of view types of intonation are located on the periphery (rhetorical question, incomplete statement) (*Figure 2*).

3. As the research has been carried out taking into account gender factor, we can conclude that though there is much likeness between male and female realisations, there are also some differences. In male speech clusterisation becomes worse if we take into account all the three parameters – fundamental frequency, intensity and duration. In this case, we cannot distinguish a single independent cluster (*Figure 1*).

The situation is different in the female speech, when consideration of all the three factors makes clusterisation more distinct (*Figure 2*).

Figure 1. Intonation Realisations Considering Fundamental Frequency, Intensity, Duration (Male Speakers)[1]



It is also worth mentioning that the most noticeable difference between male and female realisations of speech acts is revealed in emotional speech. Thus, if we view the whole complex of parameters such types of intonation as exclamation and rhetorical question do not form individual clusters in male speech and, therefore, do not possess their own prosodemic status in interference speech, being contextually conditioned variants of semantically similar types of intonation. In female realisations practically all types of intonation form individual clusters (except complete statement and statement-response), which supports the traditional opinion about greater expressiveness of female speech.

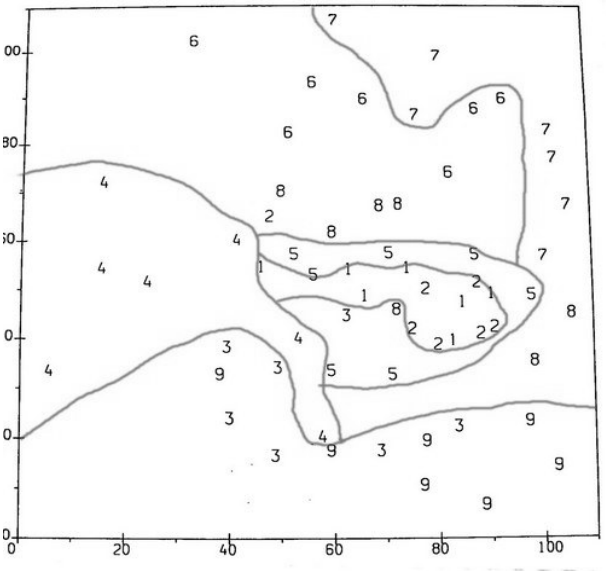
The difference in male and female realisations of special and rhetorical questions is especially evident as the realisations form two clusters (male and female) in different locations. In general female realisations form

[1] (1) Complete Statement; (2) Statement-Response; (3) General Question (transformed from a statement); (4) Exclamation; (5) Incomplete Statement; (6) Special Question; (7) Rhetorical Question; (8) Command; (9) General Question (transformed from a command).

more compact clusters which suggests lower variability of these types of intonation.

On the whole, there is much likeness in male and female realisations of speech acts; consequently there should not be special difficulties in teaching English intonation to male or female students.

Figure 2. Intonation Realisations Considering Fundamental Frequency, Intensity, Duration (Female Speakers)



4. Speaking about the information value of the acoustic parameters we can suggest that for male speech the most informative parameters are fundamental frequency and duration, whereas for female speech they are fundamental frequency and intensity. Intensity, if taken into account, allows to single out the individual cluster of statement-response and to separate the two type of general question (transformed from a statement and from a command) which is impossible if we consider fundamental frequency and duration. The other parameters (intensity for male speech and duration for female speech) fail to improve the quality of clusterisation and bring negative results. Hence, fundamental frequency

turns out to be the most “functionally loaded” parameter, which factor should be considered in teaching English intonation.

5. The analysis of phrases with the identical (and quasi-identical) syllabic structure but different basic syntactical structure shows that the last doesn’t have any particular influence on the prosodic realisation of different communicative-pragmatic types of utterance. General questions (formed from a statement and formed from a command) have similar prosodic realisation.

6. The comparative analysis of interference variants and samples of the Russian and English variants proves the existence of an intermediate, contaminated variant, which is caused by the combination of native and target languages models in the cognitive schemata of speakers.

4. Conclusion

The zone conception of intonation and the application of a specific prosodic criterion will help determine the prosodemic status of speech acts and provide the basis for the development of L2 learners’ communicative competence thus contributing to their adequate interpretation of emotional-modal connotations of utterances.

As many researches in the field notice the flaw in earlier discussions of intonation was that sentence-level intonation was analyzed and taught in terms of what intonation types typically occur in a particular language, so learners were given isolated sentences to practice. But as conversation analysis shows, natural discourse exhibits anything but "default" intonation patterns. Rather, based on the surrounding context, the speaker makes decisions about what word to stress and what attitude or intention to express. And L2 learners must be made aware early on of how stress, emphasis, contrast, and illocutionary force are expressed in the L2. Since acquiring intonation skills is closely linked to a learner’s semantic understanding, L2 teachers are urged to teach English intonation with much emphasis on communicative purposes and functions and in a socially-interactive setting.

The scope of intonation practice should be extended to include context and transactions, not only sentences. The following strategy of teaching English intonation to L2 students encompasses both drilling imitation exercises and interactive communicative practices. The training starts with demonstrating the difference in the intonation contours of the native and foreign languages through visual and auditory exercises. The positive effects of the use of visual displays of intonation for language learners are

evident. Students are told how to interpret the displays and even those who are not familiar with music notes can easily see the peculiarities of the English intonation patterns and their difference from the native ones.

The acquisition of English melodic contours begins with simple intonation patterns in limited contexts, abundant imitation exercises are accompanied by conscious interpretation and understanding of the grammar, semantic and modal meanings assigned to these patterns. The starting point is a simple nuclear tone alone, then the other parts of the intonation phrase (group) are added, such as pre-head, head and tail. Each intonation pattern is given in a short conversational situation and accompanied by the explanation of an attitude expressed (such as final, categorical or surprised and encouraging further conversation, etc.). Following the perception training activities, students would be asked to practice these utterances. Further training comprises students' responses to the given verbal context. Learners are encouraged to use one of the suggested attitudes. Alternatively, students could be offered to give their own contexts to some utterances pronounced with a certain intonation.

Gradually the context becomes broader and intonation patterns more complex. Listening activities are accompanied by careful explanation of the meaning of some nuclear tones and even whole intonation patterns in different contexts. Students are encouraged to imitate the tone, and later – to offer their own contexts and use them in their talks or conversations.

Students move to skill-based and task-based learning activities that not only offer users practice in listening comprehension but also elicit and encourage practice of specific types of interactions, language forms, sound contrasts, or nuances of meaning signalled by intonation.

Such strategy allows going beyond the sentence level and addressing the multiple levels of communicative competence: grammatical, attitudinal, discourse, sociolinguistic and phonostylistic. This way students develop their intonological ear and new cognitive schemata are being built in their minds which facilitate the process of overcoming interference phenomena in both production and perception aspects.

References

Авен, П. О., А. А. Ослон & И. Б. Мучник 1988. *Функциональное шкалирование*. М: Наука.

Grabe, E., B. S. Rosner, J. E. Garcia-Albea & X. Zhou. 2003. Perception of English Intonation by English, Spanish and Chinese listeners. *Language and Speech* 46, 4: 375–401.

Kanter, L. A., A. P. Chizhov & K. G. Guskova. 1987. A cluster-seeking technique for prosodic analysis (with special reference to Russian sentence intonation). In *Proceedings of the 11th Int. Cong. of Phonetic Sciences*, V. 4, 59–61. Tallinn: Institute of Language and Literature, Academy of Science of the Estonian S. S. R.

Odlin, T. 1996. *Language Transfer*. Cambridge: CUP.

Pike, K. L. 1945. *The intonation of American English.* Ann Arbor: University of Michigan Press.

Wennestrom, A. 2001. *The music of everyday speech: prosody and discourse analysis*. Oxford: OUP.

PART III:

APPLIED PHONETICS AND BEYOND

TO FLIP OR NOT TO FLIP? PHONETICS, PHONOLOGY AND THE FLIPPED CLASSROOM[1]

PATRICIA ASHBY

1. Introduction

In 2004, American hedge fund manager Salman Khan began tutoring his cousin in mathematics – by telephone, across several time zones. Making simultaneous use of *Yahoo*'s 'Doodle' notepad while he talked, he guided her through the intricacies of mathematical problems, writing equations and drawing diagrams to illustrate the points he was making. Ferenstein (2011) goes on to describe how other relatives and friends asked for similar tutorial help and in 2006 Khan started to make videos of his tutorials and posted them on *YouTube*. The videos were popular – everyone was able to access and use them in their own time and at their own convenience. Now publicly and freely available, however, it wasn't long before his tutoring was benefiting millions of viewers across the globe and, buoyed by their appreciation and their learning successes, in 2009 he gave up his job to develop *The Khan Academy* (see Khan Academy) full-time. His videos caught the attention of Bill Gates – whose own children used them – and today his freely accessible, virtual academy (supported by donations from both *Microsoft* and *Google*) continues to go from strength to strength. Subject areas have expanded beyond the hard

[1] This paper is an extended discussion of my UK Higher Education Academy National Teaching Fellowship research project (initially supported by a small grant from the University of Westminster, School of Social Sciences, Humanities and Languages' Learning and Teaching Enhancement Fund) and reported in Ashby (2011a) and in an earlier version of this present paper (Ashby 2011b).

sciences into history and American civics, and there are plans to add many others, including English.

During this time, two American high school chemistry teachers, Jon Bergmann and Aaron Sams, challenged by the failure of students to achieve their full potential in this subject and inspired by developments in screen-capture software, pioneered in 2007 the use of educational vodcasting (also called videocasting) to enhance the student learning experience. Demonstrable success encouraged Bergmann and Sams to persevere and perfect their technique now known as the **flipped classroom** (see University of North Colorado, also Sams 2010, Bergmann and Sams 2012). In the last five years, interest has spread around the globe – across the United States, through Canada, to Korea, Australia, New Zealand, and now the UK and Europe.

The success of this approach – this ***philosophy*** – is demonstrated by evidence not only of individual students being turned around, but whole classes, and even whole schools (such as the Clintondale High School in Michigan, USA, where flipping reduced failure rates by 33% in English, 22% in maths, and 19% in the social sciences in the space of just one semester (Clintondale High School)). Flipping enables students to learn at their own pace and in their own way. Coupled with mastery learning (Bloom 1968, Block and Anderson 1975) and structured, summative assessments (and assessment, of course, is crucially missing from the unstructured and informal context of The Khan Academy), flipping offers every student, from the strongest to the weakest, an equal opportunity to succeed. What is described here is a pilot study to gauge the effect of flipping in learning phonetics and phonology.

2. Flipping – the process

Flipping can best be defined as the asynchronous online delivery of lectures. Flipping involves switching the place of homework (private study) and lectures (formal class contact). Vodcast lectures are watched at home by students before they come to class, who then benefit from the freed-up class time to engage in hands-on work, the activity depending on the subject. They arrive in class already knowing something about the topic to be studied. They are ready to ***do*** chemistry, or phonetics, or phonology. In the case of chemistry, this might include experiments, calculations, or general lab work, for example. In phonetics and phonology, students arrive able to carry out data analysis, transcription, rule-writing, spectrography, and so on. This reverses traditional practice.

To kick-start the process, the teacher requires access to a laptop computer (with a good quality, external, plug-in microphone – inbuilt mics are not usually of sufficient quality for this purpose), appropriate software, and the PowerPoint slides normally used in the formal class-based lecture room.

Once equipped, the first decision is whether to record in class, preparing materials for use next time round, or whether to invest time in making the videos separately. My personal experience suggests that for many people, recording a live performance in class is likely to result in an end-product that has greater student-appeal – my own students found the specially prepared recording very lifeless and boring compared with their experience of me lecturing, live in the classroom. Of course, recording the lectures separately meant that it was easier for me to edit out errors and even to re-record sections if I wanted to change anything, but with hindsight, my belief is that the lack of liveliness was too high a price to pay. Students would rather see the real thing, warts and all, than a very slick, accurate, but much more sterile product that resulted from studio-based creation. Their post-trial feedback made their feelings clear. They missed the jokes, anecdotes and digressions that enter the mix spontaneously in the live lecture room. I believe there is a place for such spontaneity and that it contributes to memorability and thus to learning.

A further possibility for engendering liveliness if the lecture is not recorded in the classroom, is to record with a colleague who will act as a feed, asking questions, prompting repetition of difficult points (anticipating student queries, for example, or responding to questions asked by the tutor in order to reinforce or further clarify a point being made). This 'double act' is one that is very much appreciated by younger students who enjoy the potential for gentle levity and witty exchanges that all contribute to making concepts more memorable for them (Bergmann and Sams 2010a).

The lecture, then – the regular classroom PowerPoint slide show – is recorded using screen-capture software such as *Camtasia Sudio* (which allows for all kinds of post-editing) or *Snagit* (for quickly produced materials and messages, recorded and posted without any further editing at all). The teacher does not even feature on the screen unless he or she decides to make simultaneous use of a webcam (usually now inbuilt in the laptop) for the specific purpose of making a screen appearance. This will be an individual choice. Some like to be there in the corner of the screen throughout the lecture. Others will include themselves during the title slide only. Others again (such as myself) will make no visible appearance at all. The absence of the visible lecturer, of course, is not a bad thing. Students

are not distracted by your presence – they are free to concentrate entirely on the topic of the lecture. Your appearance – combed or uncombed hair, a stain on your tie or shirt front, a missing button (or worse), your jewellery, new sweatshirt, make-up – offers absolutely no distraction. This is exactly as it should be. The lecture is a learning process, not a staff appraisal exercise.

A graphics tablet can also be employed. Tablets mimic the typical (interactive) whiteboard facility, permitting live annotations of the kind one might add over the slides in the classroom. This enables you to write or draw on your slides during the recording, adding details or demonstrating particular points just as you would in a live lecture. (*Wacom* currently markets a wide range of such tablets, the 'Bamboo' series being particularly popular.)

For example, an early lecture in a phonetics course might focus on transcription, as per the plan in Figure 1. Reaching point 5, writing in transcription, the graphics tablet can be employed.

Figure 1. Typical lecture plan covering the concept of phonetic Transcription

Lecture: ***Transcription***

1 What is transcription

2 Types of transcription

3 Why transcribe?

4 Reading the pronouncing dictionary

5 Writing in transcription

On the slide, or in an additional pop-up box, pen strokes can be demonstrated forming symbols such as ash, **æ**, and eth, **ð**, the different descenders in palatal, retroflex and nasal consonant symbols (respectively a backward pointing or leftward hook on the left **ɲ**, a forward pointing or rightward hook on the right **ɳ** and a backward pointing or leftward hook on the right for eng, **ŋ**) or even the construct of vowel symbols such as the base shape **ɪ**, undotted I, not itself a symbol, but underlying schwi, **i** (Ashby (2011c) called lower-case i by Pullum & Ladusaw (1996), where a dot is added above), cap-i **ɪ** (also called small capital I, where a 'hat' and 'foot' are added to the base shape), and long-i, **i:** (requiring both a dot

above and a length diacritic to the right), or the diphthongs **aʊ** and **aɪ** where the first component is a printed, lower-case **a** rather than cursive, script **ɑ.** Students can actually see the shape forming as opposed to simply viewing the ready printed symbols in a text or as corrections noted on their homework.

Once the video recording is completed, it remains to post it or 'cast' it online – hence the term vodcast or videocast. These are essentially podcasts with visuals added – PowerPoint slides, for example, with an accompanying audio track.

In a UK educational establishment such as a school, college or university, publication of the completed video will usually mean uploading the file (and file format will depend on the preferred format for the system – .swf, .mp4, etc.) to a password protected Virtual Learning Environment (most commonly *Blackboard* or *Moodle)* used by the institution. Alternatively, a more widely accessible site such as *TeacherTube* can be used, or even the open-access *YouTube* favoured by Khan (Khan Academy) where your lecture can be viewed from Kochi to Kilmarnock, Belgrade to Buenos Aires, Tokyo to Trinidad. Once uploaded, the lecture is there for permanent reference – it will be watched before the class (hence the term ***pre***vodcast) either at home, outside, in the library, on the bus or train, or watched any time after class for revision purposes. The video can also be accessed in class, too, if required for questions or reference. When the vodcasts are stored as .mp4 files, students can watch them on their TVs, computers, netbooks, or mobiles, at any convenient time. Accountability is also important, however, and strategies for ensuring this are numerous, from simply taking notes and handing a copy in at class, to completion of short formative (or even summative) tests or exercises on arrival in class (a five or ten question multiple choice, for example). Commitment is also important, especially if you are flipping without the added incentive of mastery learning (see below). In the case of my own course, fully participating students had 5 bonus points added to their overall coursework grade and 87.5% (21 out of 24 who sat the final examination) participated fully and benefited from this.

In the newly liberated classroom, pedagogical strategies are designed to encourage students to take responsibility for their own learning – to give them ownership of the learning process. Underpinning everything is the now long-held belief (summarized in Dale's Cone of Experience (Dale 1969)) that students remember 90% of what they see, hear and, most importantly, ***do***... In the flipped classroom, the emphasis is definitely on doing. Flipping the classroom also recognizes that learning is a social activity (see, for example, the POGIL – Process Oriented Guided Inquiry

Learning – website). Because the classroom is no longer teacher-centred, students are no longer isolated. They have time to engage with core materials, they have the space to work in bespoke groups (the high-fliers, the strugglers, mixed-ability groups, etc.) on designated activities (which in phonetics would include learning transcription, drawing diagrams, analyzing data, etc.). The teacher, absolved from the need to stand in front and deliver a lecture, is free to move around between groups, give help or explanation where needed, challenge students at any level, and generally facilitate the learning process. Most importantly, students are now also free to learn cooperatively, to learn from each other.

So, in the context of the present study, enabled by flipping the lectures, students actually ***do*** phonology or phonetics in the classroom, rather than just hearing about it. They leave that classroom the richer for their experience and overtly confident not only in their knowledge and understanding, but also their skills. The classroom becomes a place of student-centred enquiry-based learning, rather than sterile, teacher-centred 'talk and chalk'. The classroom is about learning and not about teaching. It belongs to the students, not to the teacher. The course is delivered by means of blended learning, rather than the traditional, teacher-dominated classroom. The teacher now ***facilitates***. This is summed up by Aaron Sams who says:

> My ultimate goal as a teacher is to help students become learners who can learn for themselves and by themselves. One of the problems I was guilty of... prior to flipping my classroom around was the classroom was centred around me – I told them exactly what to learn, how to learn it, what assignments to do 'to learn it', and when to learn it, and how to prove to me that they learned it. I don't do that any more. I've changed the place in which content is delivered. Instead of standing in front of the class and delivering [...] I deliver that direct instruction now asynchronously at home through these videos that we make with *Camtasia Studio*... when kids come to class, they don't show up to learn new stuff, they show up to apply the things they learned at home and to ask me questions about the things they learned at home [...] Life is different for me because I no longer am the guy who just stands up at the front of the classroom and yacks at a student for an hour [...] Now, I walk around the class and I help kids. I'm a tutor, I'm a guide [...] I walk around and do that. I don't stand up front and teach after the kind of traditional model.
>
> (Sams 2010)

3. Flipping and mastery learning

Ideally, flipping would also eventually be coupled with mastery learning (Bergmann 2010, Bloom 1968, Block and Anderson 1975, Schools 2009). Mastery has been around for the best part of half a century, but is often overlooked in the teacher-centred classroom. Mastery hinges on learning speed and level of achievement. This necessarily requires consideration of assessment. To enable mastery, assessment should be designed to ensure that certain targets are fully achieved in the learning journey before moving on to the next point in the syllabus. In phonetics, for example, you might ensure that a student can genuinely ***read*** entries in the pronouncing dictionary before moving on to the transcription of connected speech, or that processes of connected speech have been mastered before moving on to in-depth study of coarticulation. You might ensure mastery of stress and rhythm before progressing to intonation, and so forth. If the expected grade (50% say, or 70%) is not achieved in the set task, the student then reviews the topic and tries again, ensuring that even though not all students will be at the same stage of development at the same point in time, at any given stage, nobody is lost or confused. At the same time, those who have mastered the material ahead of time are free either to progress or to exercise personal judgment and take a time management decision to work on something completely different in the phonetics class if that deadline is more pressing. A student on a combined honours programme in linguistics and English literature, for example, who is completely up to date with work in his/her phonetics module might elect to read Shakespeare's *Hamlet* during the phonetics contact hours if there is an essay deadline on the play which is more pressing than the next phonetics deadline. This is not a problem.

The sequence of learning to transcribe intonation can provide an example of mastery here. Adding intonation transcription to a text requires a number of different, progressively linked skills. Students first need to be able to apply tonality, chunking the text into intonational phrases (IPs) and within each IP, indicate sentence stress correctly. The next step is identifying tonicity – locating the nucleus or tonic syllable within each IP. Having located this prominence, an appropriate tone must be assigned. Finally, the melody of the remaining syllables must be spelled out, explicitly marking up any tail, head and pre-head. Having outlined the stages to be achieved and assigned the lowest sufficient level or grade to be targeted (50% say, in a UK educational context, or even 40% or 35% in some institutions – this is very different from countries like the USA where a much wider band of the mark scale is routinely used and where

70% or even 80% or 85% would be required to demonstrate success), students can then work on graded exercises and analyses at each of the different levels until they accomplish the prescribed grade at the final level. To progress from tonality and sentence stress (Level 1), students must achieve 50% in a summative assessment of the skill. Only then can they progress to Level 2, dealing with tonicity. Again, once they can achieve at least 50% in the summative assessment of tonicity, they progress to tone, Level 3, and likewise from Level 3 to Level 4.

Mastery Learning also, for many, provides its own incentive – students quickly realize that it is in their own best interest to achieve. The more quickly they do this, the more time they gain for other things. It is also possible within a properly designed mastery learning scheme for students to reattempt certain goals in order to improve their performance and, ultimately their grade. The bottom line is for all students to pass, but over and above that, each student determines how well they personally want to do and how much time they are willing to invest in achieving the goals they have set themselves. Of course, this requires a substantial materials bank of formative exercises and tests and of summative assessments. That in turn requires the teacher to commit preparation time to put all this in place. Once established, however, the returns far outweigh the investment.

4. Does flipping work?

4.1. The motivation

Student feedback in phonetics and phonology over the years, in my own institution and others, has highlighted antipathy to engaging with theory. In the present context, it is the theoretical content of phonetics and phonology courses that is at issue, theory normally delivered in a traditional lecture-based format. Summative assessment grades (testing knowledge of this theory) are often much lower in both these subjects than grades for hands-on practical work, and lower also than grades in other more immediately appealing subjects (sociolinguistics, for example, or language and gender). Students fail to grasp and/or remember the theory, saying there is too much of it, it is too hard, and it all takes up far too much of their time. Inspired by the flipped classroom, the *flipped lecture* aimed to remedy this.

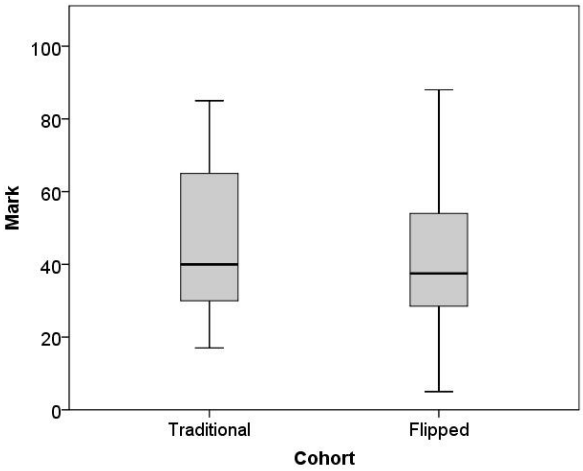
The investigation described here, focuses on final year students of phonology, studying the syllable at the University of Westminster. Phonology students were selected because they were taught as a single group. This simplified administration of the investigation by ensuring all

students had an identical learning experience and enabling a straightforward comparison of success rates across cohorts (comparing grades achieved by the ***flipped cohort*** (FC) with those of an earlier ***traditional cohort*** (TC)). Students were briefed at the beginning of the module and their agreement to participate in the trial was secured. The trial ran in weeks 9 and 10 of a 12-week teaching programme.

4.2. The evidence

Evidence of the success of flipping comes in two forms, quantitative and qualitative. To be truly successful from the student perspective, however, students must not only be seen to have 'done better' (that is, to have achieved higher marks) but they also need to be consciously aware of their enhanced grasp of the subject matter and to appreciate the process that gave rise to this achievement.

Figure 2. Comparison of grades for compulsory questions across the ***flipped*** and ***traditional*** cohorts (non-flipped topics)



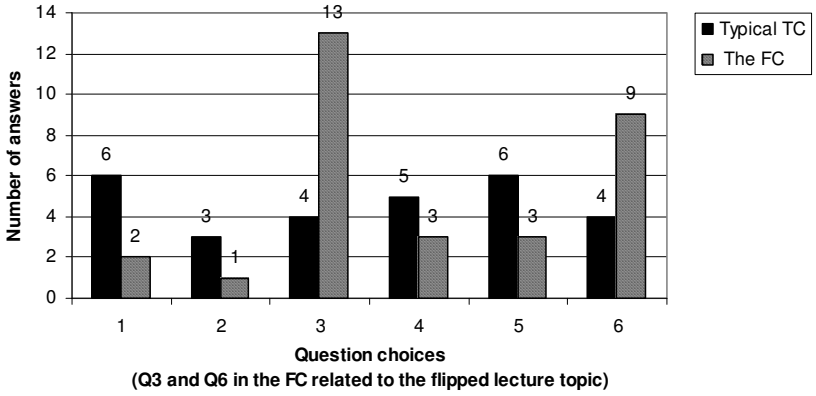
Looking first then at the quantitative evidence, the starting point was to demonstrate parity between the two cohorts (TC and FC). The final examinations taken by each cohort include two compulsory questions, the first on formalism and the second a data analysis problem. Comparison of grades for the two cohorts, TC and FC, showed no significant difference in basic ability – for Q1, $p = 0.626$ and for Q2, $p = 0.888$. The results when

pooled, as in the boxplot in Fig. 2, show very clearly the similarity of the mark range for the cohorts. Their mean marks are all but identical.

Additionally, TC grades for the free choice essay questions on the syllable showed no significant difference when compared with grades achieved for other topics ($p = 0.899$). We thus have a completely level playing field.

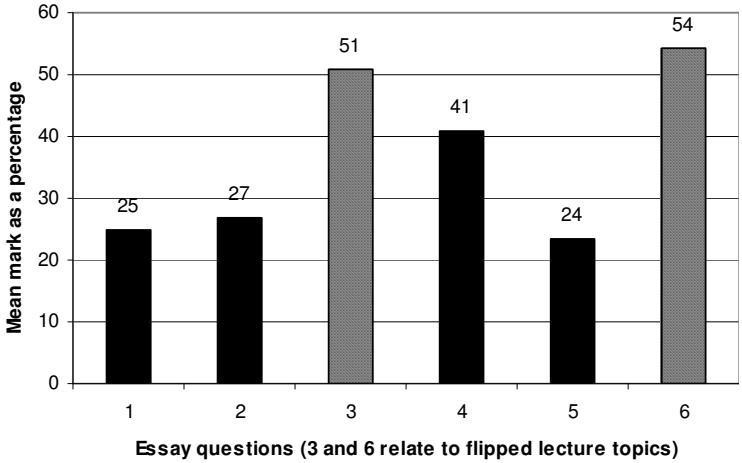
Looking more closely then at the FC results, Fig. 3 illustrates the next noticeable characteristic. In the FC cohort final examination, more students elected to answer questions from the free-choice essays on topics relating to the flipped lectures than on topics taught in the traditional classroom. In fact, 72% of all free-choice essays (22 from a total of 31 answers) were answers to questions on the flipped lecture topic. This is a highly unusual pattern for such an examination where answers are normally more evenly spread across the questions and rarely (if ever) concentrate on a single topic or area of study.

Figure 3. Distribution of answers across free-choice essay questions in the final examination



Most significant of all, however, is the level of achievement demonstrated in these answers. Anonymous papers were graded by two independent examiners and the agreed grades were moderated by an External Examiner. Looking at Fig. 4, we can see that the mean marks for the two flipped topics are 10% to 30% higher than the mean marks for answers to any other question.

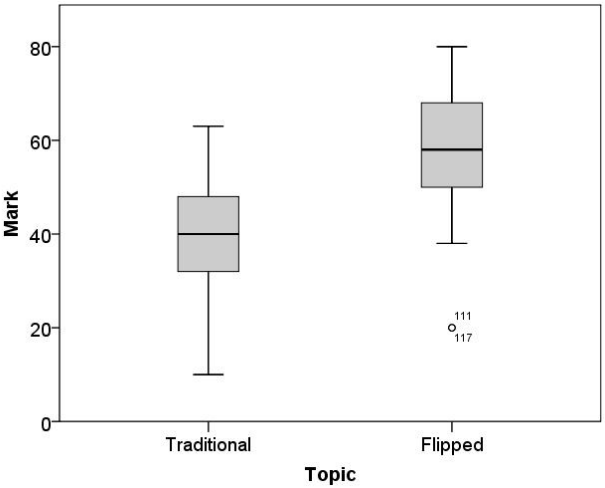
Figure 4. Comparison of mean marks for answers to free choice essay questions



Application of the Median Test (see University of Amsterdam, for example) shows that the two sets of scores differ significantly ($p = 0.0138$). In fact, given that we are dealing here with a directional hypothesis, that the answers to *flipped* questions will be better than the answers to *traditionally taught* ones, the Median Test shows this difference to be highly significant, with $p = 0.0069$. This difference is clearly visible in the boxplot in Fig. 5.

In qualitative terms, the FC students were very largely positive in their comments although there were a couple of negative issues. The first concerned fear of the unknown. After briefing, although they had agreed to participate, a mid-semester departmental student feedback survey (unrelated to the trial) revealed that 33% of them were afraid that they would be disadvantaged by flipping. They felt they would not receive 'proper lectures' or 'enough information' or that they would be unable to access the lectures because of their lack of technical prowess. Their main concern was that this would impact negatively on their degree classification.

Figure 5. Distribution of grades for the traditional and flipped answers in the final examination taken by the *flipped cohort* students.



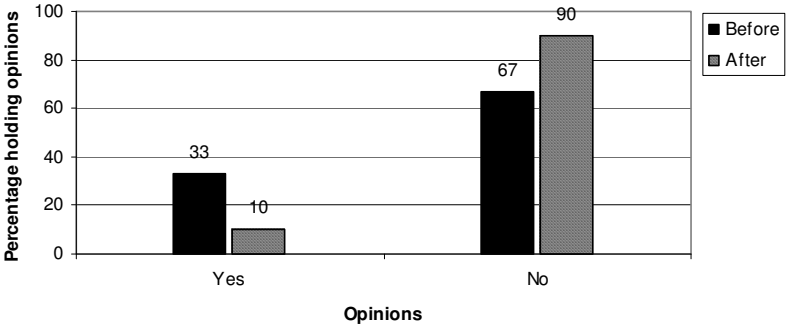
In a post-trial questionnaire, however, these fears had for the most part been dispelled. Only 10% of the respondents still claimed to feel disadvantaged afterwards (see Fig. 6) and, although all responses were anonymous, from talking informally to students my belief is that the 10% consisted of students who did not participate fully in the trial and that their feelings of being disadvantaged stemmed from the fact that they were conscious of not having the positive learning experience enjoyed by their peers. (A similar 10% strongly disagreed with the contention that flipping had enabled them to learn more than if the material had been presented in a traditional, classroom-based lecture format.)

A second criticism concerned the asynchronous nature of the lecture delivery. A small number of students – presumably those who were habitually vocal in classes – were very conscious of the fact that they could not ask questions while the lecture was ongoing. This however is simply a matter of habit – not all lecturers permit interruptions anyway, and not all lecturers take questions immediately after a lecture, either, preferring to leave these for back-up seminars or tutorials. It is also interesting to note that in spite of this feeling, very few questions were actually brought to the subsequent classes during the trial. The majority of students said that by watching sections of the lecture again, or by doing

recommended follow-up reading, they had answered their questions for themselves.

Already mentioned earlier, their final criticism concerned the sterility of the vodcast lectures. They missed, as I said, the jokes, impromptu comments, anecdotes, and exchanges that occur routinely in my regular classroom. This, I think, would be an appropriate juncture at which to consider my own role in the process and the feelings I experienced as a newcomer to vodcasting.

Figure 6. Before and after: student fears of being disadvantaged by flipping



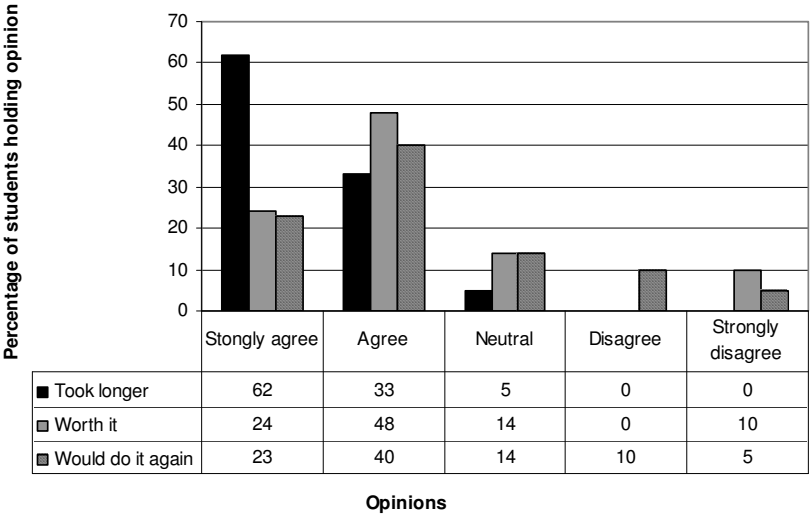
In many ways, the students' experience mirrors my own concerns – they ***saw*** what I ***felt***. As the lecturer, I found the recording process (my first experience of vodcasting) somewhat artificial. It was also quite stressful. I was working alone, trying to master a system that was new to me (Camtasia Studio) and, in the nature of the educational context, I was working against the clock under considerable time pressure. I was conscious throughout of what I felt was a need for accuracy and clarity. I felt very strongly that I should not make mistakes and then correct them. I was extremely anxious about recording a mistake that would then be there for perpetuity and even felt less confident of my own knowledge of my subject than I would ever feel in the normal classroom where words, of course, are ephemeral and the experience transient. On a more practical level, I also found it difficult to time the lecture without a live audience. All of these are things that I would expect to improve with practice. Now with hindsight, I feel that for me personally, it would probably be beneficial to make the recordings during a live session, rather than sitting, home alone, at the kitchen table. This, I believe, will differ from one

person to another. Further, while my goal at the time was to ensure that the lecture flowed seamlessly and was fully accurate, American colleagues (including the experienced and authoritative Bergmann and Sams) report that they find 'imperfect' lectures are often the best – unscripted, spontaneous performances, recorded 'warts and all' – indeed, Bergmann and Sams recommend recording the lecture as a dialogue, with two presenters rather than one (see, for example, Bergmann and Sams 2010a).

Finally, then, it remains to report on the positives. Student feedback was very encouraging in terms of their perception and awareness of the learning process and was very different from the often rather critical and negative feedback that follows traditional classes. The first point here is a blend of criticism and satisfaction, and concerns the time it took students to watch the vodcasts. Although they did not grudge the extra time involved, students were almost unanimous (with 95% agreement) in reporting that watching the lecture at home took considerably longer than the 50 minute class period. At least one student reported informally that it took at least two and a half times as long. 'Watching', of course, involved more than just sitting through the 50 minute vodcast. It included varying amounts of re-playing (95% again reported re-winding and re-playing sections one or more times), and also pausing to complete note-taking – many of the notes submitted were unusually detailed and neat compared with the scrappy jottings that take place during a live lecture (if notes are taken at all). Reactions to the time factor are presented in the chart in Fig. 7. This shows very clearly that virtually all students found the lecture took more time at home than if they simply attended a formal class, but 72% were agreed that the extra time had been worthwhile and 63% would be happy to do this mix of traditional and flipped classes again.

Following their experience of flipping, Fig. 8 shows that 62% of students (strongly) agreed that they had learnt more about this topic than about topics taught in the traditional classroom and more than half (57%) were of the opinion that more of the course should be studied this way. Their impression that they had learnt more (and in class they described how they also felt more confident about their knowledge of the topic than about other topics in the syllabus) was, of course, borne out in the final examination results described earlier – more students elected to write about the syllable (the flipped topic) and they scored higher grades on the answers they gave.

Figure 7. Summary of students' responses to flipping: the time factor

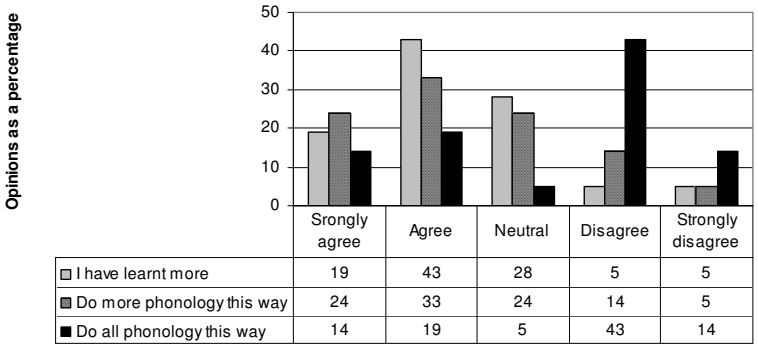


Given the amount of time students felt they had committed to flipping, it is important to know more specifically what they felt they had gained and their evaluation of the process as a learning style. Fig. 8 summarizes their conclusions.

Students' perception of the time factor, however, weighed heavily against flipping the whole module. Only about a quarter of the students felt this would be viable. That is entirely understandable, given the huge pressures on their time as final year undergraduates and given that the module as a whole was not modified to accommodate delivery in a flipped classroom. (Students were still coping with ongoing coursework assessment requirements, for example, at the same time as participating in the limited flipping trial.) It also raises the question, however, as to whether a different way of ensuring accountability (see Bergmann and Sams 2010b for one discussion of accountability) might have resulted in a different reaction. Completion of brief, formative tests on arrival in class might have involved less of the students' time outside of class. However, the crucial issue here is whether the assiduous note-taking (with time being spent producing work they felt was good enough to hand in) was itself a major contributor to the effectiveness of the learning experience. Would the same enhanced level of learning still be achieved if

accountability was demonstrated by completing of some kind of in-class formative tests? Further trials using different accountability techniques would be required to ascertain the answer to this question.

Figure 8. Summary of students' responses to flipping: gains and viability



Overall, the perceptions and reactions of the students in this UK trial reflected closely those of Bergmann and Sams's own students (Bergmann and Sams 2010c). All students were enthusiastic and understood the many benefits of flipping from the students' perspective. There was minimal negativity.

5. Conclusion

Findings tend to suggest that there is, indeed, a place for the flipped classroom in tertiary education in the UK. The UK student experience in this modest trial paralleled the experience of students at various educational levels the world over. This can only be a good thing.

The trial focused only on flipping. The next step would be to adapt class content and assessment to include mastery learning. Different countries and different educational systems will vary in how easy it is to integrate this, but inclusion in at least formative assessment should be possible in almost any context.

References

Ashby, P. 2011a. The Flipped Lecture – a Prevodcasting Trial. In *Proceedings of the Phonetics Teaching and Learning Conference – PTLC 2011*. Retrieved 6th June 2012 from http://www.phon.ucl.ac.uk/ptlc/ptlc2011/ptlc2011.php.

Ashby, P. 2011b. Phonetics and the Flipped Classroom. In *Proceedings of the 16th National Conference of the English Phonetic Society of Japan and the Second International Conference of Phoneticians of English.* Kochi: English Phonetic Society of Japan.

Ashby, P. 2011c. *Understanding Phonetics.* London: Hodder Education.

Bergmann, J. 2010. Flipped-Mastery Classroom. Learning4Mastery. Retrieved 6th June 2012 from http://www.mast.unco.edu/programs/flipped/.

Bergmann, J. & A. Sams. 2010a. The Flipped Classroom. Learning4Mastery. Retrieved 6th March 2012 from http://www.mast.unco.edu/programs/flipped/.

Bergmann, J. & A. Sams. 2010b. How to make sure students actually watch the vodcasts. Retrieved 6th March 2012 from http://www.mast.unco.edu/programs/flipped/process/accountability.php.

Bergmann, J & A. Sams. 2010c. Flipped/Mastery Educational Model: Student Impressions. Learning4Mastery. Retrieved 6th June 2012 from http://www.youtube.com/watch?v=iJrmsjdcmTY&feature=relmfu.

Bergmann, J. & A. Sams. 2012. *Flip Your Classroom: Reach Every Student in Every Class Every Day.* Alexandria, VA: ISTE/ASCD.

Blackboard. Retrieved 6th June 2012 from http://www.blackboard.com/.

Block, J. H. & L. W. Anderson. 1975. *Mastery Learning in Classroom Instruction.* New York: Macmillan Publishing Company, Inc.

Bloom, B. S. 1968. Learning for Mastery. Retrieved 6th June 2012 from http://www.eric.ed.gov/PDFS/ED053419.pdf.

Camtasia Studio. Retrieved 6th June 2012 from http://www.techsmith.com/camtasia/.

Clintondale High School. Our Story – The Flipped High School, Clintondale High School. Retrieved 6th June 2012 from http://www.flippedhighschool.com/ourstory.php.

Dale, E. 1969. *Audiovisual Methods in Teaching.* New York: Dryden Press.

Ferenstein, G. 2011. How Bill Gates' Favorite Teacher Wants to Disrupt Education. In *Fast Company.* Retrieved 6th June 2012 from http://www.fastcompany.com/1728471/change-generation-bill-gates-favorite-teacher-wants-to-disrupt-education.

Flipped Classroom Conference. 2011. Retrieved 6th June 2012 from http://www.youtube.com/watch?v=G-ej.
Khan Academy. Retrieved 6th June 2012 from http://www.khanacademy.org/about.
Moodle. Retrieved 6th June 2012 from http://moodle.org/.
POGIL. Retrieved 6th June 2012 from http://www.pogil.org/about.
Pullum, G. K. & W. A. Ladusaw. 1996. *Phonetic Symbol Guide.* Second edition. Chicago: The University of Chicago Press.
Sams, A. 2010. The Flipped Classroom. Learning4Mastery. Retrieved 16th March 2012 from http://www.youtube.com/watch?v=2H4RkudFzlc&feature=relmfu.
Schools, J. 2009. *Annotated Bibliography for Mastery Learning*. Retrieved 6th June 2012 from http://www.jimschools.com/articles/mastery_learning_annotated_bibliography.htm.
Snagit. Retrieved 6th June 2012 from www.techsmith.com/Snagit.
TeacherTube. Retrieved 6th June 2012 from http://www.teachertube.com/.
University of Amsterdam, "Median Test". Retrieved 6th June 2012 from http://www.fon.hum.uva.nl/Service/Statistics/Median_Test.html.
University of North Colorado, Vodcasting and the Flipped Classroom, Jerry Overmyer. Retrieved 6th June 2012 from http://www.mast.unco.edu/programs/flipped/.

MINIMAL PAIRS IN ENGLISH PHONETICS TEACHING

RASTISLAV ŠUŠTARŠIČ

Outline

The importance of minimal pairs in English is discussed from the point of view of English-Slovene contrastive analysis and teaching of English pronunciation.

We need to find out what phonemic contrasts there are in English, how frequently they occur and then to think about how to apply this knowledge in pronunciation classes.

The inventory of minimal pairs provided by John Higgins is briefly presented in the paper, along with a description of the main differences between the sound systems of English and Slovene, so that we can focus on those distinctive sounds that tend to be particularly problematic for Slovene students of English.

Some approaches with regard to introducing minimal pairs in pronunciation classes are suggested, as well as activities focusing on the distinctiveness of English vowels and consonants.

1. Introduction: importance of minimal pairs

Phonemic (rather than allophonic) distinctions among English (or any other) sounds seem to have a high priority in (English) pronunciation teaching, which certainly makes sense with regard to the fact that we use sounds to communicate (i.e. to *express* something) rather than to demonstrate a native-like competence in pronunciation (i.e. to *impress* someone).

It is therefore probably important to insist on the maintenance of contrasts between sounds, in particular those that "carry a high functional load" (Cruttenden 2008: 5) and in order to do this to find out what

contrasts there are, how frequently they occur at word level (i.e. what their functional load is) and (in teaching) to think about possibilities of applying this knowledge in English pronunciation classes.

An inventory of minimal pairs is available on the web page of John Higgins[1] and seems to be a good starting point for a re-consideration of the relative importance of contrasts among individual vowels and consonants, taking into account not only the functional load but also the (different or equivalent) grammatical categories of the members of each pair, and the similarity between the sounds in a pair (which usually correlates with the possibility of confusion).

Higgins defines minimal pairs as "words whose pronunciation differs at only one segment, such as *sheep* and *ship* or *lice* and *rice*" (ibid.). He has identified precisely 92,253 minimal pairs, and provides an interesting explanation of why he decided to study them in the first place (ibid.):

> (…) I once lived in a flat in the village of Etiler near Istanbul. From the living room one had a view across a green meadow down towards the steep sides of the Bosphorus, where one constantly saw passing freighters, small cruise liners and even submarines. It was one of the few places in the world where one might have said "Look, there's a sheep!" and expect to be misunderstood.

He also provides an interesting (presumably true) story about a Japanese teacher, whose mispronunciation of a geographical name had vey serious consequences (ibid.):

> Kumiko Tsuchida, a teacher of Japanese at the University of Istanbul, had a horrid time on Monday. Wishing to get to London airport to catch an evening flight back to Turkey, she instead found herself indigent on the streets of Torquay[2] well after midnight.

2. Contrasting the sound systems of English and Slovene

In teaching the pronunciation of a foreign language (in our case standard British and American English by speakers of Slovene), the best

[1] See http://myweb.tiscali.co.uk/wordscape/wordlist/.
[2] A seaside town in South West England.

starting point seems to be to present and contrast the distinctive sounds (vowels and consonants) of the two languages.

With regard to the consonantal systems, the differences are relatively small: Slovene has no dental fricatives, and everything else is either a matter of phonetic realization (e.g. places of articulation of /t d h/, manner of articulation of /r/, aspiration, glottaling and glottalization, degree of voicing), or of the phonemic vs. allophonic status of sounds in the two sound system (e.g. the phonemic status of /w ŋ/ in English and that of /ts/ in Slovene.

The vowel systems are much more problematic due to the distinctiveness of different types of short and long (and also qualitatively different) vowels in English.

The contrasts that seem to be particularly problematic for Slovene students of English are the following:

- voiced/weak vs. voiceless/strong obstruents in final position (*race/raise, lap/lab, rate/raid, leak/league*)
- contrast between dental fricatives and (dental) plosives (*three/tree*)
- contrasts among similar vowels (in particular *sheep/ship, lend/land, fussed/fast,* and *pool/pull*).

The functional load (in terms of the number of minimal pairs) for most of these contrasts is quite high; there are (according to Higgins, ibid.) 466 minimal pairs contrasting the vowels of *sheep/ship,* 305 with those of *lend/land,* 170 with those of *fussed/fast,* and 128 for the initial consonants of *three/tree*. However, there are only 18 minimal pairs for the contrast *pool/pull.* With regard to the contrasts between final weak and strong obstruents, we can perhaps simply look at the overall number of minimal pairs with homorganic plosives, affricates and fricatives, although it is of course possible to extract only those examples in which these consonants occur in final position. As we can see from the overall numbers of minimal pairs for the obstruents below, the functional load is generally high, except for the dental and post-alveolar fricatives:

612 for p/b
867 for t/d
444 for k/g
106 for the affricates in *batch/badge*
153 for f/v
11 for the fricatives in *wreath/wreathe*

281 for s/z
9 for the fricatives in *Confucian/confusion*

3. Approaches to minimal pairs in pronunciation teaching

In order to make sure that students perceive differences between different phonemes (in particular vowels), a good starting point may be to develop activities based on non-words rather than existing English words. This ensures that students do not try and guess the word on the basis of context, disregarding the actual articulation of the word in question. This may be the main reason for the 'identification' exercises in pronunciation courses like that of A. C. Gimson (Gimson 1975), in which in addition to valid English words also some non-words (called 'nonsense words' by Gimson) are used for transcription tasks.

When dealing with actual English words, we can start by contrasting words with similar, yet distinctive sounds, such as the vowels of *beat-bit, bet-bat, but-bart, pot-port, pull-pool*, and then move on to using such words in connected speech.

When I was an English student, we were presented with sentences containing several words with the same vowel, such as the following (author unknown to me):

We need tea for three, please - for Jean, Steve and me.
Fit six thin bricks into this big tin lid.
My head got better when I went to bed at ten instead of eleven.
'It's hard to park such a large car in the dark', Arthur remarked.
The cook took a good look at the pudding and put some sugar in it.

Notice that minimal pairs do occur in these sentences; however, they are not examples of contrastive vowels but consonants, such as those of *thin – tin* etc., so that such sentences may be more useful for teaching consonantal rather than vowel phonemes.

In our textbook (Collins et al. 2002), we have a number of sentences in which similar vowels occur within sentences, but the contrasted words are not minimal pairs, e.g.:

I think Keats is pretty difficult for these kids.
I never carry cash when I travel.
The gang's come armed for trouble.

One of my favourite activities which aims to enhance the students' awareness of the distinctiveness of sounds is what I call 'the phonemic journey', in which students 'travel' from one lexical item (the starting point) to another by way of adding, deleting and/or replacing one phoneme in each step of the way, as for example when going from *death* to *life* or the other way round, e.g.:

death – debt – let – light – life

A couple of other examples with such antonyms would be:

take – bake – break – brick – bring

strong – strung – stung – tongue – rung – wrong – rock – lock – leak – weak

This can easily be extended to the rhyme in poetry (which, incidentally, provides one of the interesting links between language and literature classes, or more precisely between phonetics and 'poetics'). We can ask students to provide the missing words in a poem; very often these words form minimal pairs with those occurring in the lines between, e.g.:

Whose woods these are I think I know.
His house is in the village ----
He will not see me stopping here
To watch his woods fill up with ----[3].

The minimal pairs in this case are *know-though, know-snow,* but not, of course, *though-snow.*

Notice that also the whole point of the so-called 'spoonerisms' (named after Reverend W.A. Spooner) is playing with minimal pairs, as in the famous example below[4]:

[3] Robert Frost, *Stopping By Woods on a Snowy Evening*.
[4] See http://www.youtube.com/ watch?v=gmOTpIVxji8.

'Sir, you have tasted two whole worms, you have hissed all my mystery lectures and been caught fighting a liar in the courtyard. You will leave Oxford by the next town drain'.

In order to show the distinctiveness of words with different phonemes in interaction that pretends to be 'realistic' rather than fictional, one can try and find contexts in which the response of the student with one of the suggestions will tell us if the word has been identified correctly. I came across examples of this kind a long time ago at a UCL Summer Course of English Phonetics in London (the author(s) of these examples are unknown to me):

I'm going to SLEEP/SLIP.
- Good night. See you in the morning.
- Hold on. I'll help you.

Please CALM/COME down.
- I can't help it. I'm extremely upset.
- I'll be down as soon as I can.

While it is difficult enough to find words that can be used in the same syntactic function (to begin with, they must be of the same grammatical category), it is even more challenging to think of contexts where such words could really be confused. We can see this also in different exercises available on the Internet, which are provided for listening comprehension, e.g. the following sentences on a web page[5]:

My neighbors soothed/sued me often.
My friend comes from a very loyal/royal family.
My friends had a lot of wines/vines in their basement.
They wondered when they were going to suffer/supper.
All the students saw the three/free men and applauded.
The rest/rust of the car was too much for Fred to work on.

[5] See http://www.docstoc.com/docs/135204268/ENG-3205-Minimal-Pairs-Exercise.

It seems clear that in such sentences one of the words would usually fit the context better than the other; furthermore, it seems that it is hard enough to find any minimal pair for a particular sentence, let alone finding words in which the contrasted phonemes would be similar and/or problematic enough to practice. Thus in the last sentence, the vowels of *rest – rust* could hardly be confused by anyone, as they differ both in terms of place of articulation (front versus central) and degree of opening (close-mid to open-mid versus open-mid to open).

Concerning the grammatical categories, Higgins points out that "(…) two nouns, such as *beer* and *pier*, are much more confusable than a noun and a preposition, such as *frog* and *from*" (ibid.)

It seems that (precisely due to the lack of convincing examples of potential confusion) we are more likely to come across such examples in jokes and movies than in real life, as for example in the famous video clip ridiculing a German speaker's pronunciation of *thinking* as *sinking* (available on YouTube[6]).

Almost half a century ago, Trim (1965) was obviously aware that we should approach the issue of contrastive sounds in a light-hearted unrealistic manner, providing amusing sentences (with accompanying illustrations), e.g. *The zoologist wonders about bugs*. or *The botanist wonders about bogs*.

Ideally, we should try and find examples of confusion for those phonemes (either vowels or consonants) that are relevant for our students. Thus the example above might be useful for German speakers of English, but not Slovene, who replace dental fricatives with dental or alveolar plosives rather than alveolar fricatives. Thus Slovene students would typically pronounce *three* as *tree,* and *there* as *dare,* etc. I have managed to find an amusing example of this confusion in a popular movie *My Cousin Vinny*, in which Vinny pronounces the word *youths* as *Utes.*

With regard to the very common (and notorious) neutralization of pairs of similar English vowel phonemes, there is of course the famous Italian speaker's complaint in a hotel about requiring *a sheet (on the bed), a piece (of cake)* and *a fork (on the table),* but the problem with this joke is that typically non-native speakers of either Romance or Slavic origin will use a

[6] See http://www.youtube.com/ watch?v=gmOTpIVxji8.

close front vowel instead of a mid-close centralized one rather than the other way round, and that the vowel of *fork* is rather unlikely to be pronounced as a central unrounded vowel; thus in the latter case, the joke around the title character in the movie *Meet the Fockers* is (at least from the point of view of standard American rather than British pronunciation) a more appropriate example of confusion, even for native speakers of English.

Finally, when dealing with phonemic contrasts, we should not ignore the seemingly less important allophonic variations. It has often been pointed out, for example, that native speakers of (standard British or American pronunciation) distinguish between strong and weak (voiceless and voiced) plosives more on the basis of ('non-distinctive') aspiration of the former than on the 'distinctive' features of voicing and/or strength of articulation. I have tried to find examples of confusion caused by the lack of aspiration of /p t k/, but have managed to find only an example of the opposite, that is of a native speaker of English (actor Rowan Atkinson) replacing a weak unaspirated initial /g/ with a strong aspirated /k/: the example comes from a scene in the movie *Keeping Mum,* in which Atkinson, in the role of an absent-minded Reverend Walter Goodman, delivers a lecture about the *cod's* (i.e. *God's) mysterious ways.*

References

Collins, B., R. Šuštaršič & S. Komar. 2002. *Present-day English Pronunciation: A Guide for Slovene Students.* Ljubljana: English Department, Faculty of Arts, University of Ljubljana.

Cruttenden, A. 2008. *Gimson's Pronunciation of English.* Seventh edition. London: Hodder Education.

Gimson, A. C. 1975. *A Practical Course of English Pronunciation: A Perceptual Approach*. London: Edward Arnold.

Trim, J. 1965. *English Pronunciation Illustrated*. Cambridge: CUP.

BEGINNINGS, ENDINGS, AND THE IN-BETWEENS: PROSODIC SIGNALS OF DISCOURSE TOPIC IN ENGLISH AND SERBIAN[1]

TATJANA PAUNOVIĆ

Outline

The study presented in this chapter focuses on prosodic cues used to signal discourse topic structure (topic beginning, continuation, and ending) in a reading task performed by two groups of participants – L1 speakers of Serbian, who were also EFL learners, and L1 speakers of English (Great Britain). The analysis included F0/pitch, intensity, and duration measured at intonation unit boundaries (left and right edges), first peak/onset, and nuclear accent syllable, as well as overall intonation unit pitch range and intensity. The findings point to some similarities and differences between English and Serbian, and suggest that some, but not all of the EFL students' problems could be attributed to L1 prosodic transfer.

1. Introduction

The discourse-structuring function of prosody has been extensively researched in English and various other languages. In Serbian, however, studies of prosody that would go beyond the level of individual utterances to take account of discourse structure are much fewer. Apart from the research on lexical pitch accent, sometimes including its interaction with intonation in utterances (e.g. lnkelas & Zec 1988; Ivić & Lehiste 1996; Kašić 2000, 2012; Lehiste & Ivić 1963, 1986; Zsiga & Zec 2012, *inter alia*), or those aiming to formulate a more up-to-date phonological

[1] The research presented here was part of the project *Languages and cultures in time and space* (No. 178002), supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

representation of the Serbian prosodic system (Godjevac 2000; Marković 2011; Smiljanić 2003), there are very few studies that deal with the discourse-related functions of prosody (cf. Bogetić 2010; Polovina & Panić 2011 on conversational prosody).

Even less attention has been paid to Serbian learners of English. There are very few studies focusing on EFL learners' use of prosody (Marković 2011; Paunović & Savić 2009), and fewer still that would investigate the prosodic signals of topical structure, although this aspect of prosody is very important for both speech comprehension and speaking proficiency. It can be affected by various factors, including L1 transfer, which, as repeatedly found in cross-linguistic research, can occur at the levels of both phonological organization and phonetic realization. Furthermore, since many prosodic aspects of speech are of gradient rather than categorical nature (Grice & Baumann 2007), speakers of different languages can perceive the continua, e.g. of pitch movement, in different ways. It is, therefore, very important to investigate the discourse-organizing function of prosody cross-linguistically.

2. Discourse topic prosody research

The notion of *discourse topic* has been defined in different ways, ranging from Brown and Yule's (1983) simple statement that topic is 'what a piece of discourse is about' (Brown & Yule 1983: 69), to Chafe's (1994) explanation that in coherent discourse the topic links the "related events, states, and referents" (Chafe 1994: 121) and represents "the totality of information that is semiactive at one time" (Chafe 1994: 128).

Chafe (1994, 1997) describes discourse organisation as a hierarchical structure, in which the minimal unit of thought organization is 'intonational unit', or 'a momentary focus of consciousness'. These are grouped together into a 'center of interest', or a 'superfocus of consciousness', signalled by a falling pitch "perceived as sentence-final prosody". Such 'centers' are grouped into discourse topics, which "constitute more stable units of mental representation", and which can contain several 'subtopics' (Chafe 1994: 139-144, 1997: 395-396).

In addition to its content, topic is often signalled formally, by discourse markers or 'signposts' (Chafe 1997: 390), of which prosody is especially important. Topic beginning is typically signalled by 'heightened volume, pitch, and tempo', while their end is signalled by 'petering out', often also 'creaky voice', 'an iconic lengthening' of the last word, and significant pausing before a new topic is introduced' (Chafe 1997: 396). Particularly important are the ends of intonation units, where, supported by amplitude,

duration, and voice quality, 'terminal pitch contours' signal either 'forward-looking' or 'backward-looking', that is, they either anticipate something to come, or signal that 'something has arrived at closure' (Chafe 1997: 399).

Although Chafe's description is based on 'analysis by introspection' and hypotheses in Xu's terms (2011: 88), abundant empirical research actually supports this kind of description. Hirschberg and Pierrehumbert (1986) showed that relationships in discourse structure are signalled by systematic variations of pitch range, so that major unit boundaries are marked by its largest increases (cf. Grosz & Hirschberg 1992; Grosz, Hirschberg & Nakatani 1994; Hirschberg & Grosz 1992).

Introduction of a new discourse topic is signalled by high pitch (Lehiste 1975; Brown *et al.* 1980), 'resetting the baseline' (Vaissiere 2005; Mennen 2007) and an initial rise (Vaissiere 2005); when the topic is changed, the onset F0 is high, with the maximal peak F0 considerably raised, while the final F0 of the preceding utterance is low, sometimes accompanied by laryngealisation, decreased intensity, and a narrowing of the pitch range (Nakajima & Allen 1993; Wichmann 2000).

Finality, including the end of discourse topic, has been found to be signalled by F0 fall, lowered F0 contour (Venditti & Swerts 1996; Vaissiere 2005), boundary tones lower than for continuation (Swerts & Geluykens 1994), compressing the pitch range (Brown *et al.* 1980; Lehiste 1975), the duration of pauses (Lehiste 1979; Swerts & Geluykens 1994; Hirschberg & Nakatani 1996; Grosz & Hirschberg 1992; Grosz, Hirschberg & Nakatani 1994), segmental lengthening (Vaissiere 2005), and downstep and downdrift (Hirschberg & Pierrehumbert 1986; Vaissiere 2005). Wichmann (2000), too, points out the tendency toward 'supradeclination', as a gradual descent in pitch from the beginning to the end of a discourse unit.

Studies of prosodic cues signalling discourse structure look into a growing scope of languages, and many of them involve cross-linguistic comparisons. However, comparisons that would include Serbian are quite rare (Godjevac 2000; Marković 2011).

One point often highlighted in cross-linguistic comparisons, and particularly important for L2 learning, is that the phonetic implementation may be as relevant – and as difficult for students – as the phonological form. This point was particularly stressed, for instance, by Mennen (2006, 2007), with respect to the use of pitch range and tone alignment (the temporal realization of tones, i.e. the precise timing of a peak or a valley). This can be language-specific, and, although not phonologically relevant, it can be an important cue in both speech perception and the evaluation of L2 speakers' proficiency (Mennen 2007). Marković (2011: 248), on the

other hand, warns that the line between phonetic and phonological differences can be difficult to draw, and that more cross-linguistic research is needed. Regarding prosodic signals of discourse structure, Mennen (2007) lists several problems identified with L2 learners of different L1 backgrounds, such as signalling intonation group boundaries, discourse topic introduction, continuation and termination, and particularly the reset after a final boundary.

3. Present study

3.1. Aims and methodology

The aim of this study was to investigate the acoustic prosodic cues used to signal discourse topic initiation, continuation, and termination in English and Serbian, in the context of EFL learning.

It is important to highlight two points about the research design and methodology. Firstly, several authors have pointed out (Brown *et al.* 1980: 27; Hirschberg 1993: 90; Swerts & Geluykens 1994: 23) that discourse prosody research runs the risk of being circular unless the link of prosodic cues to discourse segments is investigated with the discourse structure established previously through an independent analysis. To achieve this, some studies have relied on specific models of discourse structure (Chafe 1994, 1997; Grosz & Sidner 1986), while others have combined perception and production investigations (cf. Xu 2011). In this respect, the present study can be described as experimental in that the texts used as read-aloud materials were designed and prepared specifically for this study, and were segmented into topics, subtopics, and intonation units independently, by three different listeners. All the three were lecturers in the English Department of the Faculty of Philosophy, University of Niš, with ten or more years of teaching experience in EFL, linguistics, and applied linguistics. In other words, the listeners were language professionals in Wichmann's (2006) sense, and, at the same time, L1 speakers of Serbian. Therefore, the segmentation into discourse topics and subtopics was performed independently of the acoustic analysis, though from a communicative rather than theoretical point of view. Furthermore, the segmentation by the three listeners showed almost perfect inter-reliability. The two intonation units whose status within their respective topics was judged differently by two out of three listeners were excluded from the analysis, and the acoustic measurements were performed for only those discourse segments whose status all the three participants agreed on.

Secondly, the study focused on one phonological unit – intonation unit (intonational phrase), with four potentially relevant domains: the left and right edges, the intonation unit 'onset', i.e. the first pitch peak, and the syllable carrying the nuclear (pitch) accent (right-most 'phrase accent' in Serbian). In this sense, the study was based on a hypothesis about the phonological units relevant for signalling discourse structure in Xu's (2011) sense, but did not include a more thorough phonological description, or other phonological units and prominence relationships proposed for English and Serbian, e.g. prosodic/phonological word, or lexical pitch accent, since they are more relevant for other phenomena, such as utterance-level information structure, than for discourse topic. The investigation focused primarily on the acoustic cues, and a discussion of their phonological status was not part of the study.

3.2. Participants and materials

The study aimed to observe and compare the performance of a) native speakers of English (Great Britain) reading a text in English, b) native speakers of Serbian reading a text in Serbian, and c) the same L1 Serbian speakers, EFL students, reading the English text.

Therefore, one group of participants (EFL) consisted of four male EFL students (approximately at the B2+ level of proficiency by CEFR), 2nd year students of the English department, Faculty of Philosophy, University of Niš, all native speakers of Serbian, specifically, the urban variety spoken in Niš, the largest town in South-East Serbia. The other group (NS) included four male L1 speakers of British English, from Southern England. All the participants were of approximately the same age (20-26).

Two texts were used in the read-aloud tasks: one in English, read by both NS and EFL participants, and the other in Serbian, read only by the Serbian participants. As segmented by the three listeners in the preliminary analysis, the text in English (230 words) consisted of 52 intonation units (IU), grouped into 5 discourse topics. The topics contained 7–17 intonation units (IU). The Serbian text (260 words) was *not* a translation of the English text, but was designed to match the English text in terms of its discourse structure and information structure. It consisted of 53 IUs grouped into 5 discourse topics, which contained 9–15 IUs each. Both texts contained portions of narration and portions of dialogue/ conversation. Both contained examples of different types of utterance-level focus, for diversity and naturalness, and were matched as closely as possible for the syntactic structures used (declarative sentences, yes/no questions, Wh-questions, intonation questions, question tags, vocatives,

interjections etc.). The texts were also matched for genre (anecdote), level of formality (semi-formal to casual everyday speech), and general theme (an encounter with strangers in a public place).

Although reading has been criticized in favour of spontaneous speech in prosody research, we opted for this kind of task because it is very important for EFL students. Namely, while spontaneous conversation data can doubtlessly offer a better insight into the way prosodic cues are used to signal intended meanings and govern interaction, meaning-negotiation, and repair, for L2 learners the skill of reading is as important, because, in addition to oral fluency, it also involves the ability to interpret (correctly or otherwise) the meanings intended by the text writer. Therefore, if, as Wichman (2006: 2) states, reading involves a special kind of 'professional' skill, "a learned skill, possibly subject to cultural conventions", for L2 students this skill is especially important, because oral reading fluency has been shown to be a good predictor of better comprehension and better overall linguistic competence. After all, even if factors such as motivation, identity, and group dynamics can influence students' interpretation of the emotional, attitudinal, and interactional prosodic functions, at least their basic comprehension of the text can be expected to be signalled by an appropriate use of prosodic cues to identify discourse topic structure.

Therefore, we can sum up the research questions the study focuses on in this way:

1. Within each of the participant groups (NS and EFL) – is there any correlation or systematic difference between the analysed acoustic cues and the identified topic-structure positions – topic beginning, topic continuation and topic end?

2. Between the participant groups – are there any differences in the way the observed acoustic cues were used in these three positions by different groups of speakers, specifically between:

    a) the English text read by the NS group and the Serbian text by the EFL group?
    b) the NS group and EFL group reading the English text?
    c) the EFL group reading the Serbian text and reading the English text?

3.3. Analysis

The procedures of data analysis included making relevant acoustic measurements (Praat 5.2.04, Boersma & Weenink 2010), and a statistical analysis of thus obtained data through a number of procedures (SPSS v. 13). The acoustic measurements included:

- pitch level – F0 maximum (max), minimum (min), mean;
- pitch range, in Hz and semitones (ST);
- intensity maximum, minimum, mean;
- duration (of pauses, word syllables), in milliseconds (ms).

The measurements were taken at four points selected as relevant domains for topic structure: the left and right edges of the intonation unit, the first pitch peak (onset), and the nuclear accent (phrase accent). Pitch range was measured for whole topics, for each IU, and for the post-nuclear/ accented IU part.

The statistical analysis of the data obtained through acoustic measurements included the following:

1) For within-group comparisons (Research question 1) one-way ANOVA comparisons of means, and Krusak-Wallace's test as the nonparametric counterpart. Correlation procedures were applied to check whether the data would show a connection between a specific parameter/cue value and the position in the discourse topic.

2) For between-group comparisons (i.e. between-text/reading comparisons, Research question 2), the statistical procedures included Independent samples T-test.

4. Results and discussion

4.1. NS data

In native speakers' reading (NS group), the discourse topic structure was signalled by several acoustic cues, mainly in line with previous research findings. Firstly, this was observable in the raw data. **Table 1** shows some of the average NS values (F0, intensity, pause duration, pitch range) for IUs relative to their position in the discourse topic.

Compared to topic-medial and topic-final IUs, in topic initial IUs the left edge consistently showed the highest F0 values – maximum, minimum and mean, as well as the highest intensity. The left edge of topic-medial and topic final IUs showed a regular decrease of F0 values (maximum, minimum and mean), as well as a decrease in intensity. The first peak

(onset) of topic-initial IUs also had the highest intensity and F0 values (maximum, minimum, mean).

Table 1. Average measurement values for NS speakers (English) – F0/pitch, intensity, pitch range (in Hz and ST) and pause duration after the IU in three structural positions – topic initial, medial and final

	IU left edge (beginning)				**IU first peak (onset)**			
	F0 max Hz	F0 min Hz	F0 mean Hz	Int. mean dB	F0 max Hz	F0 min Hz	F0 mean Hz	Int. mean dB
topic initial	170.0	133.0	140.0	70.0	206.2	156.8	182.6	72.0
topic medial	140.6	126.4	134.4	65.6	173.9	134.9	154.1	67.3
topic final	122.0	115.0	118.3	64.3	188.8	129.5	159.5	67.2

	IU right edge (end)				**pause after IU**
	F0 max Hz	F0 min Hz	F0 mean Hz	Int. min dB	duration sec.
topic initial	118.7	91.0	100.3	59.3	.124
topic medial	113.6	88.7	100.2	55.1	.272
topic final	104.0	79.8	89.8	55.2	.436

	IU span			**IU nucleus**			
	pitch range Hz	pitch range ST	Int. max dB	F0 max Hz	pitch range ST	postN F0 min	postN range ST
topic initial	102.8	11.8	70.4	154.0	6.6	107.0	0
topic medial	91.6	12.1	68.3	145.1	6.5	95.2	0.8
topic final	105.2	14.5	67.8	133.6	5.2	79.8	3.7

A relatively regular downstep tendency could be observed from the beginning towards the end of the topic, particularly at the IU left edge, but also in the F0 values of the IUs' right edge, where all F0 values showed a regular decrease, including the gradual lowering of the baseline. The regular downstep was partly overridden at the first peak, where F0 minimum did decrease gradually from topic initial to topic-medial and topic-final positions, but the F0 maximum was higher in topic-final IUs

than in topic-medial ones, suggesting that the first peak might be a domain more relevant for signalling prominence and information structure, which can override the otherwise regular downtrend as a discourse-topic signal.

The end of the topic was marked by a drop in F0 and intensity at the right edge of the topic-final IU. Also, it was signalled by lowering, rather than overall narrowing the pitch range in topic-final IUs. In these IUs, the minimum F0 values were indeed the lowest, but the overall pitch range was the widest, wider even than that in topic initial IUs, suggesting that other factors, such as prominence or information structure influence the overall pitch range in topic-final IUs. However, the post-nuclear part of the topic-final IUs showed a wider range than in topic-initial and topic-medial IU, suggesting that, after the most prominent syllable, F0 dropped and the pitch range was lowered and compressed as a topic-ending signal.

The most important signals of topic finality, however, were the significantly increased pause duration and creak (laryngealization), used by virtually all NS speakers at the right edges of all topic-final IUs.

The statistical analysis supported these observations, too. Statistically significant correlations (nonparametric Spearman), and/or significant differences (ANOVA, Kruskal-Wallis) between IUs relative to their position in the discourse topic were found for the following acoustic cues.

A (negative) correlation (r=.030 at the .05 level) was found between the IU position in the topic and the maximum pitch at the beginning of the IU (beginning pitch maximum dropped as the TU position moved from 1. initial, via 2. medial, to 3. final, as the database was coded), supporting the observed downstepping trend. A statistically significant difference (=.026) was found between IUs' first peak (onset) F0 maximum with respect to their position within the topic. Also, a correlation was found between the IU's position in the topic and the minimum F0 value of the whole IU (r=.030, .05 level). The mean intensity in the IU decreased, too, for topic medial, and especially topic final IUs (r=.032, at the .05 level).

Significant differences between IUs with respect to their position within a topic were found (Kruskal-Wallis) for mean intensity at the left edge (.012), while a negative correlation (r=.006, .01 level) was observed between the intensity minimum at the first peak and the topic position of the IU, as well as a somewhat weaker correlation (r=.031 at the .05 level) for the first peak intensity mean. What these correlations cannot specify, but can be observed in the comparison of the raw data, is that intensity was increased in topic-initial IUs, at the left edge and the first peak, and when an initial IU was observed as a whole. Initial IUs' overall intensity minimum, maximum and mean values were regularly higher than in topic-medial and topic-final IUs. However, no regular decrease in intensity

could be observed throughout a discourse topic, since the values for topic-medial and topic-final IUs were rather close together, while only topic-initial intensity values stood out clearly.

The strongest correlation (r=.007 at the .001 level) was observed for the use of creak to signal the end of the topic, which proved to be a cue used most regularly by the NS readers. This was confirmed by statistically significant differences (Kruskal-Wallis, .006), as well as by non-parametric tests. That the duration of the pause was regularly increased to signal the end of a discourse topic was also supported by an observed correlation (r=.024 at the .05 level), confirmed by non-parametric tests.

Therefore, it can be said that the results of our NS group comply in most details with the findings of numerous previous research studies.

4.2. EFL English data

In reading the English text, EFL students used the investigated acoustic cues much less consistently and systematically. The only correlations between the IU topic position and an acoustic cue value were found for the use of creak at the end of topic-final for IUs (.005, at the .01 level) and the duration of the topic-final pause (.01, significant at the .01 level). These two were also the only cues found to produce statistically significant differences between IUs with respect to their topic position by a non-parametric test (Kruskal-Wallis) – the creak significance .006 at the .05 level, the duration of the pause significance .042 at the .05 level.

Raw data comparison shows just a slightly different picture (**Table 2**). For instance, topic-initial IU values for the F0 maximum, minimum and mean were somewhat higher than in topic-medial and topic final IUs, but topic-final IUs had a higher F0 maximum, and only slightly lower minimum and mean, while the intensity at the left edge of IUs showed almost no difference between topic-medial and topic-final IUs. Similarly, the first peak (onset) average F0 values (maximum, minimum, and mean) were indeed the highest, but topic-medial and topic-final IUs had almost the same average values. Therefore, although the beginning of a new topic could be said to have been signalled by a higher initial pitch, not even raw data suggest a steady and gradual downstep throughout a topic.

As for the overall pitch range in IUs, a tendency to use a wider pitch range in topic initial than in topic-medial IUs could be observed, especially with respect to the highest F0 produced. The narrowest range was observed in topic-medial IUs, as well as the lowest average F0 minimum values. In topic-final IUs, the pitch range was overall not significantly narrower, although in their post-nuclear part, the wider pitch-

range than in topic-medial and topic-initial IUs did suggest a tendency to compress the range and to further lower the F0.

At the right edge of IUs, the difference between topic-initial and topic-medial IUs was generally rather small, but the end of topic-final IUs indeed showed the lowest average values for both pitch (maximum, minimum, mean) and intensity, suggesting that EFL speakers did use pitch and intensity lowering, in addition to increased pause duration, as topic-ending signals. However, the statistical analysis did not support any of these observations.

Table 2. Average measurement values for EFL speakers (English) text – F0/pitch, intensity, pitch range (in Hz and ST) and pause duration after the IU in three structural positions – topic initial, medial and final

	IU left edge (beginning)				**IU first peak (onset)**			
	F0 max Hz	F0 min Hz	F0 mean Hz	Int. mean dB	F0 max Hz	F0 min Hz	F0 mean Hz	Int. mean dB
topic initial	154.7	136.0	144.0	70.0	252.0	146.0	196.7	68.3
topic medial	129.6	120.8	125.2	66.9	181.7	137.4	161.5	69.8
topic final	134.3	112.3	122.3	66.7	183.2	134.8	163.6	69.7

	IU right edge (end)				**pause after IU**
	F0 max Hz	F0 min Hz	F0 mean Hz	Int. min dB	duration sec.
topic initial	104.5	95.5	100.0	62.0	.154
topic medial	115.8	105.1	108.9	64.2	.328
topic final	90.8	82.3	86.5	60.8	.606

	IU span			**IU nucleus**			
	pitch range Hz	pitch range ST	Int. max dB	F0 max Hz	pitch range ST	postN F0 min	postN range ST
topic initial	133.2	13.6	73.4	174.8	8.0	95.5	0
topic medial	78.9	9.3	76.0	147.9	4.1	106.4	1.7
topic final	92.4	11.7	76.4	127.2	4.1	82.3	3.3

4.3. EFL Serbian data

While reading the text in Serbian, EFL speakers used acoustic cues in a clear and systematic way, as shown by several correlations, confirmed by non-parametric tests, as well as by significant differences between IUs relative to their position in the topic.

The beginning of a discourse topic was clearly signalled by a higher pitch. F0 values (maximum, minimum, and mean) at the left edge of topic-initial IUs were regularly higher than in topic-medial and topic-final IUs (F0 max negative correlation =.000 at the .01 level, F0 min =.001, .01 level, F0 mean =.016, .05 level; significant differences for F0 mean =.021, min =.047 and max =.004). The mean intensity values were also higher at the beginning of topic-initial IUs (negative correlation =.017, .05 level).

The first peak (onset) F0 values also showed correlations with the IU topic position – F0 maximum (negative correlation =.004, .01 level) and F0 minimum (=.019, .05 level), also supported by the significant differences (Kruskal-Wallis test) between IUs relative to their topic position in the values of the first peak pitch span in ST (significant difference =.026).

The end of a topic was regularly signalled by a drop in the pitch values at the right edge of the IUs – the lowest F0 values were found in topic-final IUs for the F0 maximum (r=.026, .05 level), F0 minimum (r=.014, .05 level) and F0 mean pitch (r=.020, .05 level). Topic finality was also signalled by a decrease in the intensity values at the right edge of topic-final IUs – mean intensity (=.002, .01 level), and minimum intensity (=.042, .05 level).

The cues used the most regularly were the creak at the end of a topic-final IU (r=.000 at the .01 level) and increased pause duration(r=.000 at the .01 level), both correlations further supported by the statistically significant differences (pause duration =.002, creak =.001).

The pitch range was found to correlate with the topic position of IUs (Pearson correlation r=.029 at the .05 level), both for F0 maximum (=.004, .01 level) and F0 minimum (=.001, .01 level). These correlations were further supported by the statistically significant differences (Kruskal-Wallis test) found for IU pitch range in HZ =.016, IU F0 maximum =.006 and F0 minimum =.012.

Raw data, presented in **Table 3**, reveal a similar picture, but suggest some more specific observations. The beginning of a discourse topic was consistently signalled by higher pitch values at the beginning and the first peak, by a wider and higher overall IU pitch range, as well as by higher intensity. The end of the discourse topic was signalled by a drop in the

right-edge pitch values (maximum, minimum and mean), a drop in intensity, and, most regularly, by a prolonged pause and a creak.

Table 3. Average measurement values for EFL speakers (Serbian text) – F0/pitch, intensity, pitch range (in Hz and ST) and pause duration after the IU in three structural positions – topic initial, medial and final

	IU left edge (beginning)				**IU first peak (onset)**			
	F0 max Hz	F0 min Hz	F0 mean Hz	Int. mean dB	F0 max Hz	F0 min Hz	F0 mean Hz	Int. mean dB
topic initial	208.0	155.0	175.0	71.3	252.0	146.0	196.7	68.3
topic medial	115.4	103.2	118.8	65.4	181.7	137.4	161.5	69.8
topic final	95.3	91.0	94.3	64.6	183.2	134.8	163.6	69.7

	IU right edge (end)				**pause after IU**
	F0 max Hz	F0 min Hz	F0 mean Hz	Int. min dB	duration sec.
topic initial	118.0	114.6	116.6	62.0	.080
topic medial	113.5	104.1	109.2	57.7	.226
topic final	81.4	75.8	78.6	55.4	.598

	IU span			**IU right-most accented syll.**			
	pitch range Hz	pitch range ST	Int. max dB	F0 max Hz	pitch range ST	postN F0 min	postN range ST
topic initial	148.4	15.1	78.6	179.4	2.8	131.0	1.9
topic medial	76.8	9.9	76.7	130.4	2.6	110.8	1.9
topic final	77.2	10.3	73.2	109.6	2.5	76.2	2.0

However, raw data suggest that in the Serbian text the left edge and first peak were the domain more important for signalling discourse topic beginning, since left-edge values between the topic medial and topic-final IUs were closer together, while only the topic-initial values stood out clearly. Conversely, the right edge of the IU was the domain more relevant for signalling topic finality, because the right-edge values for topic-initial and topic-medial IUs were closer together, and only the topic-final IU

values clearly stood apart. Therefore, the raw data do not show a continuous and gradual downstepping trend of all values from topic beginning to topic end.

As for the pitch range, the values (in Hz and semitones) suggest that it would be more precise to say that the pitch range was widened and heightened to signal the beginning of a topic, rather than stating that it grew narrower or lower towards the end of the discourse topic. Namely, the pitch range in topic-medial IUs was even narrower and a bit lower than in topic-final IUs, showing no gradual lowering towards the end of the topic, while topic-initial IUs had the widest and highest pitch range.

Therefore, we could say that our findings are in some respects in line with the observations made in previous research. Regarding finality, Polovina and Panić (2011: 386) observed that topic ending is signalled by a low-falling tone, and that the nuclear tone of the topic-final utterance is lower compared to other nuclear tones in the same topic, which was evident in our data, too. Godjevac (2000) points out that the sentence-final position in declarative utterances is characterized by distinctive intonational shape, "a highly reduced pitch range with the pitch very close to the speaker's bare line" (Godjevac 2000: 134), i.e. 'laryngealization' (Lehiste & Ivić (1986: 186, in Godjevac 2000: 134). Indeed, the laryngealization or the creak, together with the increased pause duration, was the acoustic signal most regularly and consistently used by the participants to signal discourse topic ending, even though the data concerning the pitch range compression offered no clear patterns.

Godjevac proposes that the phonological representation of this final lowering be "a L- phrase accent" (Godjevac 2000: 137), a property of 'intonational phrase' as a higher-level phonological constituent, manifested through "lowering the ceiling of the pitch range at the edge of the constituent", which, in neutral prosodic conditions, is the right-most constituent. The influence of the final position on a peak is to make it "lower than it would have been if it were not in the final position (Godjevac 2000: 138). Therefore, this higher-level final-position phrase accent affects the shape of lexical prosodic accents, and modifies them through pitch range manipulation (Godjevac 2000: 146). Three types of prosodic realisation are possible in this final position – in a broad-focus utterance, after narrow focus prominence, or when the word itself is prosodically prominent, i.e. focused. Only in the case of focal prominence would the pitch range of the final constituent be expanded, while it would be compressed in both other situations (Godjevac 2000: 157).

This could be very relevant for our data. However, our study aimed to 'capture' broad trends in the use of the acoustic cues of pitch, intensity and

duration within the topic structure of a narrative discourse, and did not include the possible effects of lower-level prominence phenomena, such as lexical pitch accent in Serbian, or the interaction with utterance-level information structure (type of focus, given vs. new information, contrast). Therefore, further analyses are necessary to investigate these interactions. Similarly, a peculiar downtrend pattern in longer utterances, a certain "pleating effect", i.e. a "partial reset of the pitch range at constituent boundaries" (Godjevac 2000: 141) in which each peak in the reset can be as high or even slightly higher than the preceding one (Godjevac 2000: 145), was not clearly observed in our data, probably because our materials did not contain only broad focus declaratives.

As for signalling initiality, Godjevac (2000) points out that "both the sentence initial position and the discourse initial position [...] have the highest H target of all", but the initial position is "set off from the rest by the relatively higher pitch target regardless of its syntactic status" (Godjevac 2000: 132). Our data also showed that the highest pitch (and the widest pitch range) characterized the IU left-edge, including its first peak.

4.4. Between-group/text comparisons

Comparing the ways in which acoustic cues were used by the two groups of speakers in the two texts, several observations can be made.

4.4.1. Firstly, the NS English text was compared to the EFL Serbian text, i.e. each text read by its respective native speakers.

Topic beginning was signalled in both English and Serbian by high pitch and intensity at the left edge of the IU, the highest pitch and intensity on the first peak, and the higher pitch range in topic-initial IUs. Notably, the pitch range of topic-initial IUs was also wider and not only higher in Serbian, though not in English.

Topic end was signalled in both English and Serbian by a significant lowering of the pitch, which was in Serbian accompanied by an observable decrease in intensity, too. The most regularly used cues of topic finality in both languages were a creak (laryngealization) at the right IU edge, and increased pause duration.

The main difference between the two languages occurred with respect to downstep. In English, a clear and gradual downtrend throughout the topic, from the beginning to the end, was observable at both the left and the right edges of IUs, and it even, to a certain extent, included the first peak, particularly its minimum pitch, indicating a steady decrease in the baseline throughout the topic. In Serbian, however, our data suggested a different pattern – the left IU edge and the first peak clearly signalled

discourse topic beginning, but did not show a steady downtrend towards the end of the topic. Similarly, the IU right edge was clearly the domain of topic finality, but did not show a steady downtrend through the topic.

Another difference suggested by our data is linked to the pitch range manipulation. In English, the pitch range was lowered and compressed in the post-nuclear part of the topic-final IU. In Serbian, however, a more accurate observation based on our data would be that the pitch range was widened and heightened to signal the beginning of a topic. The pitch range can be said to be lowered rather than narrowed at the end of a topic.

4.4.2. Comparing the NS and EFL groups reading the English text, the only two cues that showed an obvious similarity in the performance of both groups was the use of topic-final creak, and the increased duration of pauses to signal topic ending. Apart from these, the comparison of raw data, although not supported by statistics, suggested that both groups used higher pitch at the beginning of topic-initial IUs, as compared to topic-medial IUs. The end of discourse topic was signalled by a lowered pitch at the right-edge of topic-final IUs, accompanied in the EFL group's performance by an observable drop in intensity, too. All these were, in fact, the features shared by the English and Serbian texts when read by their respective native speakers.

The most striking observation, however, was that EFL speakers did not use acoustic cues to signal topic structure in a way that would be regular, consistent and systematic enough for correlations or statistically significant differences to emerge in the analysis. The fact that some patterns could still be discerned in the raw data suggests a process observed in investigating other aspects of EFL students' performance, too, namely, a sort of 'destabilization' in the students' interlanguage phonology, so that certain features are sometimes used appropriately, and sometimes not, without consistency. EFL students' performance comprised some L2 and some L1 effects, while some were unlike either of the languages.

An illustration of this can be found in a three-way comparison of one of the cues, the pitch at the left edge of the IU relative to its topic-structure position. **Figure 1** shows left-edge F0 maximum plotted (from left to right in each picture) for topic-initial, topic-medial, and topic-final IUs as found in 1) NS English text, 2) EFL Serbian text, and 3) EFL English text. While in both English and Serbian texts read by their respective native speakers F0 maximum at the left edge of the IU was the highest in topic-initial IUs, lower in topic-medial IUs, and the lowest in topic-final IUs, in their reading of the English text EFL students did not follow this pattern.

Figure 1. Left-edge F0 max plotted (from left to right in each picture) for topic-initial, topic-medial and topic-final IUs as found in 1. NS English text, 2. EFL Serbian text, and 3. EFL English text.

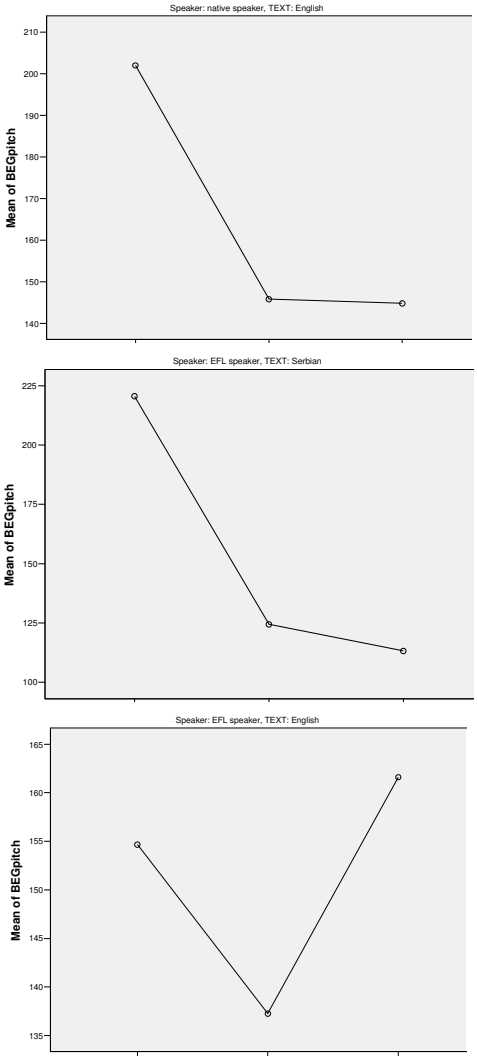
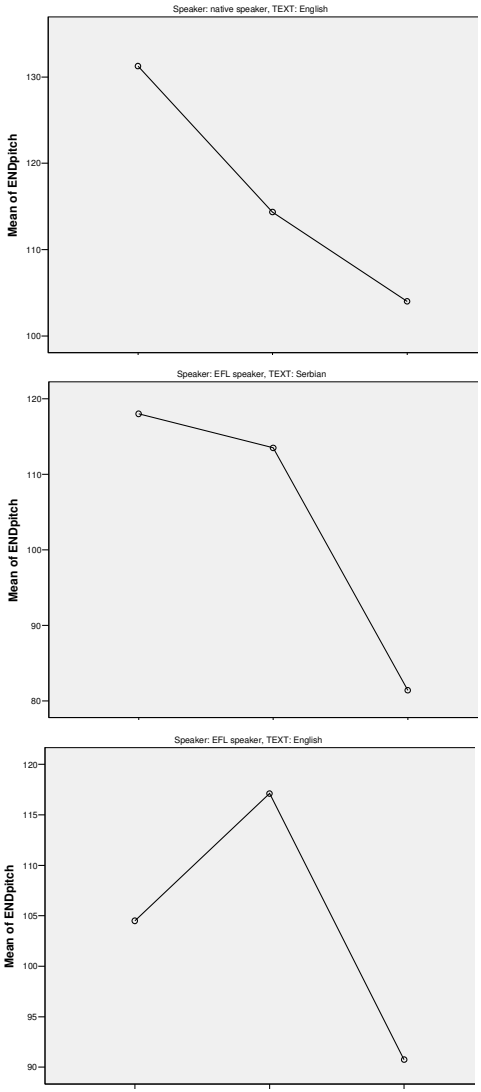


Figure 2. F0 maximum at the right edge of the IU, plotted (from left to right in each picture) for topic initial, medial and final IUs, as produced by 1. NS 2. EFL in Serbian, and 3. EFL in English



4.4.3. The comparison of the EFL group's performance in reading the English and Serbian texts pointed to some features that may explain the perception of inappropriately used topic-related prosodic cues in L2. For instance, **Figure 2.** shows F0 maximum at the right edge of the IU plotted (from left to right in each picture) for topic initial, medial and final IUs, as produced by 1. NS 2. EFL in Serbian, and 3. EFL in English.

NS speakers' maximum F0 at the right edge showed a steady, albeit not striking, gradual lowering from topic-initial, through topic-medial, to topic-final positions. In Serbian, as discussed above, instead of a gradual lowering of right-edge maximum F0, only the topic-final IUs showed a significant drop, while topic medial and topic initial IU values were similar and notably higher. Based on this, we proposed that in Serbian, unlike English, the left IU edge may be the domain more relevant for signalling topic beginning, while the right edge is more relevant for signalling topic ending, which would explain the absence of a clear and gradual downtrend throughout the topic.

However, EFL students reading the English text failed to produce a clear pattern that would reflect either English or Serbian trends, and used the cue appropriately only in topic-final IUs. Whether this could still be ascribed to L1 transfer, if we assume that the right-edge F0 is used as a relevant signal only in topic-final IUs, where the students did use it to signal topic ending, while they disregarded it as irrelevant in topic initial and topic-medial IUs, remains to be further explored empirically. What these examples show clearly, though, is that EFL students can have problems with the use of acoustic cues in L2 even if the given cues are used in very similar ways in L1 and L2, as in the example illustrated in Figure 1, apart from the problems with acoustic cues used in L1 and L2 in different ways, as in the example illustrated in Figure 2.

Another observation based on a three-way comparison between NS and EFL group's Serbian and English texts concerns the manipulation of the pitch range. The common impression of EFL teachers that Serbian students do not use an appropriately wide pitch range in English, which is, again impressionistically, attributed to L1 transfer, was actually not supported by our data. As can be seen in **Table 4,** the pitch range used by EFL students in English, as well as in their Serbian text, was barely narrower than the range used by the NS group. In fact, the widest average pitch range value occurred in EFL students' reading of the Serbian text, in the topic-initial IUs. Compared to the NS group, the EFL students produced a *wider* pitch range in topic-initial IUs in their English text, too,

while their pitch range was a bit narrower only in topic-medial IUs, in both their English and Serbian texts.

This falls in line with Marković's (2011: 248) remark that "native speakers of Serbian seem to use a repertoire of low and high pitch on a par with the native speakers, regardless of the perceived transfer, i.e. a strong foreign accent", and that the reasons for the "perceived 'higher pitch' in English, especially BE" still need to be investigated. Whether EFL students' pitch range manipulation could be explained by L1 transfer, if we assume that in Serbian pitch range is manipulated primarily to signal topic beginning, in topic-initial IUs, while it is not a used as a relevant cue in topic-medial or final positions, remains to be further explored.

Table 4. Pitch range (in ST) – average values for the NS group English text, EFL group English text, and EFL group Serbian text, in topic-initial, topic-medial and topic-final IUs

	IU pitch range in ST		
	NS English	EFL English	EFL Serbian
topic initial	11.8	13.6	**15.1**
topic medial	12.1	9.3	9.9
topic final	**14.5**	11.7	10.3

Our analyses did not include other possible functions of pitch range manipulation, for instance, to signal information structure, focus, or contrast, which might explain the widest pitch range used by the NS group in topic-final IUs. It is possible that acoustic cues used to signal information structure are not the same in Serbian and English. These questions were not within the scope of this study, but our data clearly show that much further research is needed into the prosodic functions and their acoustic signals in Serbian, particularly in comparison to English.

5. Conclusion

The main limitation of this study is the one common to much phonetic research, namely, a small a number of participants. Therefore, although the study was extensive, in that the eight participants produced a corpus of over 650 'cases' to analyze for 66 variables, with three sets of data to compare within and between groups, the findings cannot be generalized to the use of prosodic cues in Serbian, or to EFL students in general,

particularly if we bear in mind that L1 is an abstract construct, and that transfer derives from the students' specific native variety of L1.

Still, our findings did highlight the fact, stressed by Grice and Baumann (2007: 31), too, that a close comparison of L1 and L2 is essential for L2 learning and teaching, particularly when it comes to prosody. Our study showed that otherwise relatively proficient EFL students, for whom a thorough comprehension of a very simple anecdote in L2 cannot be an issue, still had problems using prosodic cues while reading it, even in the relatively simple task of identifying and signalling discourse topic structure. They had problems even with some cues that were apparently similar in L1 and L2, which especially highlights two facts. Firstly, L1 transfer is a very complex phenomenon, which requires a close investigation of both L1 and L2. Secondly, students can be expected to master the use of prosodic cues to signal discourse-related effects only through instruction and practice, and these should focus on the areas identified as relevant by comparative L1-L2 empirical research. Awareness raising about both L1 and L2, a powerful tool in many areas of L2 learning, may be particularly important in acquiring L2 prosody, where few other tools have proved efficient enough.

References

Boersma, P. & D. Weenink. 2010. Praat: Doing phonetics by computer (Version 5.1.30) [Computer program]. Retrieved 20th November 2010 from http://www.praat.org/.

Bogetić, K. 2010. Napomene o ulozi prozodije u diskursu konverzacije – nalazi iz govora u interakciji na engleskom i srpskom jeziku. *Nasleđe*, Kragujevac 7 (16): 205–220.

Brown, G. & G. Yule. 1983. *Discourse Analysis.* Cambridge: CUP.

Brown, G., K. Currie & J. Kenworthy. 1980. *Questions of Intonation.* Baltimore: University Park Press.

Chafe, W. 1994. *Discourse, Consciousness, and Time: The flow and displacement of conscious experience in speaking and writing.* Chicago: The University of Chicago Press.

Chafe, W. 1997. Interplay of Syntax and Prosody in the Expression of Thoughts. *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure*, edited by M. L. Juge & J. L. Moxley, 389–401. Berkeley, CA: Berkeley Linguistics Society.

Godjevac, S. 2000. *Intonation, Word Order, and Focus Projection in Serbo-Croatian*. Unpublished PhD thesis, Ohio State University.

Grice, M. & S. Bauman. 2007. An introduction to intonation – functions and models. In *Non-Native Prosody: Phonetic Description and Teaching Practice,* edited by J. Trouvain and U. Gut, 25–51. Berlin/ New York: Mouton de Gruyter.

Grosz, B. J. & J. Hirschberg. 1992. Some intonational characteristics of discourse structure. In *Proceedings of the Second International Conference on Spoken Language Processing* (ICSLP-92), edited by J. Ohala et al, 429–432. Banff, Edmonton, Canada: Personal Publishing Ltd.

Grosz, B. J. & C. L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12: 175–204.

Grosz, B. J., J. Hirschberg & C. H. Nakatani. 1994. A study of intonation and discourse structure in directions. In *Working papers of the Workshop on the Integration of Natural Language and Speech Processing,* 124–131. Menlo Park, CA: American Association for Artificial Intelligence.

Hirschberg, J. 1993. Studies of Intonation and Discourse. *ESCA Workshop on Prosody,* Lund, Sweden, September 27–29, 1993. ISCA Archive. Retrieved 18th January 2011 from www.isca-speech.org/archive.

Hirschberg, J. & B. Grosz. 1992. Intonational features of local and global discourse structure. In *Proceedings of the Speech and Natural Language Workshop on Spoken Language Systems*, February 23–26, Harriman, New York, *ACM International Conference on Human Language Technology Research*, 441–446. Harriman, NY: DARPA, Morgan Kaufmann.

Hirschberg, J. & C. H. Nakatani. 1996. A prosodic analysis of discourse segments in directiongiving monologues. In *Proceedings of the Thirty-fourth Annual Meeting of the Association for Computational Linguistics,* Santa Cruz, 286–293. New York: Association for Computational Linguistics.

Hirschberg, J. & J. Pierrehumbert. 1986. The intonational structuring of discourse. *Proceedings of the Twenty-fourth Annual Meeting of the Association for Computational Linguistics,* 134–144. New York: Association for Computational Linguistics.

Inkelas, S. & D. Zec. 1988. Serbo-Croatian pitch accent: the interaction of tone, stress and intonation. *Language* 64: 2227–2248.

Ivić, P. & I. Lehiste. 1996. *Prozodija reči i rečenice u srpskohrvatskom jeziku* (prevod Lj. Subotić). Novi Sad/Sremski Karlovci: Izdavačka knjižarnica Zorana Stojanovića.

Kašić, Z. N. 2000. The function of suprasegmental elements in speech expression. *Beogradska defektološka škola* 2–3: 113–124.

Kašić, Z. N. 2012. Absolute end of an expression as the phonetic position. *Beogradska defektološka škola* 2: 309–324.

Lehiste, I. 1975. The phonetic structure of paragraphs. In *Structure and Process in Speech Perception, Proceedings of the Symposium on Dynamic Aspects of Speech Perception, I. P. O,* Eindhoven, August 4–6, Vol. 11, edited by A. Cohen & S. G. Nooteboom, 195–206. Berlin, Heidelberg: Springer Verlag.

—. 1979. Perception of sentence and paragraph boundaries. In *Frontiers of Speech Research*, edited by B. Lindblom & S. Oehman, 191–201. London: Academic Press.

Lehiste, I. and P. Ivić. 1963. Accent in Serbo-Croatian: An experimental study. *Michigan Slavic Materials 4.* Ann Arbor: Universtity of Michigan, Department of Slavic Languages and Literatures.

—. 1986. *Word and Sentence Prosody in Serbocroatian.* Cambridge, MA: MIT Press.

Marković, M. 2011. Acquiring second language prosody: Fundamental frequency. In *Proceedings of the First International Conference on English Studies, English language and Anglophone literatures today - ELALT, Novi Sad, 19th March 2011*, edited by I. Đurić Paunović and M. Marković, 238–249. Novi Sad: Filozofski fakultet.

Mennen, I. 2007. Phonological and phonetic influences in non-native Intonation. In *Non-Native Prosody: Phonetic Description and Teaching Practice*, edited by J. Trouvain and U. Gut, 53–76. Berlin & New York: Mouton de Gruyter.

—. 2006. Phonetic and phonological influences in non-native intonation: An overview for language teachers. *QMUC Speech Science Research Centre Working Paper WP9 (2006)*, series editors J. M. Scobbie, I. Mennen, and J. Watson.

Nakajima, S. & J. F. Allen. 1993. A Study on Prosody and Discourse Structure in Cooperative Dialogues. TRAINS, Technical Note 93–2, September 1993.

Paunović, T. & M. Savić. 2009. Discourse Intonation – Making it Work. In *ELOPE* V, 1–2. *As You Write It: Issues in Literature, Language, and Translation in the Context of Europe in the 21st Century*, edited by S. Komar and U. Mozetic, 57–75. Ljubljana: Slovene Association for the Study of English and Department of English, Faculty of Arts, University of Ljubljana.

Polovina, V. & N. Panić. 2011. Suprasegmentna obeležja u diskursu TV debata. In *Simpozij Obdobja30, Meddisciplinarnost v slovenistiki/ Interdisciplinarity in Slovene Studies,* edited by S. Kranjc, 383–389. Ljubljana: Filozofska fakulteta.

Smiljanić, R. 2003. Lexical and pragmatic effects on pitch range and low tone alignment in two dialects of Serbian and Croatian. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 39, 1, 520–539. Chicago: Chicago Linguistic Society.

Swerts, M. & R. Geluykens. 1994. Prosody as a marker of information flow in spoken dicsourse. *Language and Speech* 37(1): 21–43.

Vaissière, J. 2005. Perception of intonation. In *Handbook of Speech Perception*, edited by D. B. Pisoni & R. E. Remez, 236–263. Oxford: Blackwell.

Venditti, J. J. and M. Swertsz. 1996. Intonational cues to discourse structure in Japanese. In *Proceedings of the Fourth International Conference on Spoken Language Processing,* ICSLP 96*,* 725–728. Philadelphia: University of Delaware & Alfred I. duPont Institute.

Wichmann, A. 2000. *Intonation in Text and Discourse: Beginnings, Middles and Ends (Studies in Language and Linguistics)*. Harlow, England: Longman.

Wichmann, A. 2006. Prosody and discourse, a diachronic approach. In *Proceedings of Discourse-Prosody Interface symposium – IDP05*, Aix-en-Provence, Seridisc, Belgique, edited by C. Auran, R. Bertrand, C. Chanet, A. Colas, A. Di Cristo, C. Portes, A. Regnier & M. Vion. CD-ROM.

Xu, Y. 2011. Speech prosody: a methodological review. *Journal of Speech Sciences* 1(1): 85–115.

Zsiga, E. & D. Zec. 2012. Contextual Evidence for the Representation of Pitch Accents in Standard Serbian. *Language and Speech* 56(1) 69–104.

BRITISH OR AMERICAN PRONUNCIATION?

SNEZHINA DIMITROVA

Outline

This article compares the students' pronunciation preferences with their spoken performance, based on analyses of the students' recordings. Forty-seven recordings were analysed both auditorily as well as with reference to signal waveforms and pitch traces in Praat, in order to establish how consistent the Bulgarian tertiary-level learners were in their use of the most well-known salient segmental and suprasegmental features of the pronunciation model of their choice Vowel quality of words from the LOT and BATH lexical sets, along with rhoticity and t-voicing, position of lexical stress and variable individual word pronunciations such as *clark* /klɑːk/- /klɜːk/ were among the most prominent traits that students used inconsistently when trying to imitate the British RP or the American GA accent.

1. Introduction

In a paper presented at BIMEP 2010, we discussed pronunciation models taught to university students of English in Bulgaria (Dimitrova and Chernogorova 2012). It was based on an investigation which involved, among other things, a survey in which we asked 47 freshers to name the accent on which they wanted to model their own pronunciation, and then recorded them reading a short diagnostic passage. We found that native-speaker models, in particular Received Pronunciation (RP) and General American (GA) were preferred to non-native ones by the students who took part in the study.

Bulgarian students are by no means unique with respect to their preference for native-speaker accents of English. A number of studies have looked at students' attitudes towards different varieties of the language. As noted by Llurda (2009: 123), though results are somewhat

diverse, there is one common feature among learners from L1 backgrounds as diverse as Italian, Danish, or Austrian German, and it is their preference for RP accented British English over their local accents. In addition, it is very often the case that, given a choice between native-speaker accents, a big majority of students opt for standard British rather than American pronunciation: this was the case with our students in Sofia, it has also been reported for Danish (Ladegaard and Sachdev 2006), for Polish - in several papers presented at the Poznan Linguistic Meetings (see Dziubalska-Kolaczyk and Przedlacka 2005), as well as for learners from a number of other first language backgrounds.

While encouraging our students from the very start to choose a native-speaker target, we also require that they be consistent: for example, that they try not to allow any rhoticity to intervene in an otherwise British-style accent. This insistence on consistency is a problem which comes in addition to the problems already faced by the average Bulgarian learner of English: the pronunciation of the dental fricative sounds, final devoicing, rhythm and weak forms, etc. It is therefore important that we are aware of the different aspects of the problem of consistency. The present study aims at investigating this problem in a more systematic way, mainly focusing on the segmental aspects of teaching a British vs. an American accent to tertiary level students.

2. The investigation

2.1. Introduction

In a survey which we conducted some time ago, we asked forty-seven freshers at the Department of English and American Studies of Sofia University to name the accent on which they wanted to model their own pronunciation. The students had to complete the sentence "I would like to speak English like…". We also asked the students to read and record a short diagnostic passage.

The results from the survey were as follows:

- a little less than two-thirds (27 students, or 57.5%) said that they wanted to sound British,
- seven students (15%) replied that they wanted to speak like an American,
- two students (4%) wanted to sound Australian, and
- about a quarter (11 students, or 23.5%) replied that they hadn't decided yet.

The analyses reported in this paper look at the recordings of the students from the first two and the last of the above groups. We focused on features which are known to best distinguish a standard British from a standard American accent. A brief note on terminology is in order here. In spite of various objections that have been raised against the terms "Received Pronunciation (RP)" and "General American (GA)" when talking about "standard pronunciation" on the two sides of the Atlantic, we have chosen to use them as shorthand for the two most widespread pronunciation models that are usually taught to learners of English around the globe.

The most salient features which distinguish the two accents are well-known, and have been described in works dedicated to the study of variety such as Wells (1982, 1997), Kortmann and Upton (2008), Schneider (2008), and also in phonetics and phonology textbooks such as Cruttenden (2008) and Collins and Mees (2008), among others.

2.2. Diagnostic passage

Dimitrova (2003: 118) gives a short diagnostic passage designed to elicit some of the most important of these characteristics (see Appendix 1). This was the passage which the students who took part in the investigation had to record. The salient features which were the focus of the analysis are discussed below in the order in which they occur in the passage.

1. Yod dropping: omission of the palatal glide, especially after alveolar consonants, is typical of many (though not all) GA speakers. In the word *new* yod dropping was expected to occur in the speech of students who had chosen "standard American" as their pronunciation model.

2. Words from the LOT lexical set have the PALM vowel in GA. The diagnostic passage predicts 17 occurrences of this feature. Two of these were judged to have relatively low predictive power, since they can be reduced, namely, the "short o" vowels in *everybody* and *nobody*. The stressed vowel in *wandered* was later also discarded from the analysis because a number of students were unsure of the correct pronunciation, and may thus have been influenced by the spelling of the word.

A limited subset of LOT words has /ɔ:/ in GA. Of these, the word *gone* occurred in our passage, so for it, both /gɔ:n/ and /gɑ:n/ were expected (compare Wells (2008: 347) who gives the following poll results for GA: /gɔ:n/ 76%, /gɑ:n/ 24%).

It is well-known that, as a group, the vowels of the standard lexical sets LOT, CLOTH, PALM, THOUGHT are historically unstable, so in American English there is quite a lot of variability with respect to low

back vowels. Accordingly, for words such as *on, office,* and *thought* (although the latter word was eventually excluded from the final analysis), variants with both /ɑː/ and /ɔː/ were judged acceptable for students aiming at an American accent.

3. Words from the BATH lexical set have the TRAP vowel, usually before voiceless fricatives, or before a nasal followed by another consonant. Four such words occurred in the dianostic passage, three of which were included in the final analysis, while the realisation of the final vowel in the word *photograph* was discarded because it can have a front realization in RP as well (given by both Wells (2008) and Roach et al. (2011).

4. Rhoticity: this feature included the pronunciation in GA of *r* both in post-vocalic position (e.g., *girl*), as well as after consonants and word-finally (e.g., *car park*), and in words of the NEAR, SQUARE, CURE lexical sets where RP has centring diphthongs. A few instances of word-final /r/ were not considered diagnostic because they could have been used as r-links in an otherwise RP-style accent. Overall, 15 occurrences of rhoticity in a GA-style accent were expected. (Pronunciations of *clark* with a mid-central rhotic vowel were considered under 6. below.)

5. Voicing and/or tapping of /t/ in intervocalic position within words and across word boundaries: we expected 12 occurrences of this feature in the reading of students who wanted to "sound American", e.g., *pretty*, *thought it*.

6. The diagnostic passage also included a number of lexical items whose pronunciation differs unpredictably in British and American English, namely, *clerk* and *leisure*, as well as some words in which the suffix vowels are different in the two standard accents. Of the latter, only *hostile* was included in the analysis.

The realisation of the vowel in words belonging to the GOAT lexical set was originally meant to be used as a diagnostic feature. However, the vowel is notoriously difficult for Bulgarian learners of English. Only a couple of the students aiming at an RP-style accent produced GOAT words in the passage (e.g., *told*, *nobody*) with a central starting point. The rest had a rounded back starting point which, in our experience, is largely due to the way in which they were taught to pronounce the respective words. The GOAT words were excluded from the final analysis since their diagnostic power to distinguish between RP and GA-style accents was judged to be low in the case of Bulgarian learners.

The use of schwa instead of "short i" in unstressed syllables (e.g., *office*) is a feature on which pronunciation dictionaries sometimes tend to disagree. Thus, the Longman Pronunciation Dictionary gives /ˈɑːf əs/ as

the recommended GA variant (Wells 2008: 559), whereas the Cambridge English Pronouncing Dictionary has /ˈɑː.fɪs/ instead (Roach et al. 2011: 348). Although the diagnostic passage was meant to elicit this feature, we eventually decided not to include it in our analysis.

Finally, some features of a GA-style accent which the diagnostic passage was not designed to elicit, were t-deletion from *nt*-clusters, the use of the same vowel in words such as *marry – merry – Mary*, production of "dark l" in all contexts, closer realisation of the STRUT vowel, etc. Last but not least, in terms of differences in setting, intonation and rhythm, which Collins and Mees (2008: 156–7) include among the salient features distinguishing British from American accents, our recordings in general sounded "Bulgarian-accented", with obvious mother tongue interference. Suprasegmental features were not analysed for the diagnostic purposes of this investigation.

The analysis thus focused on the following salient characteristics:

- LOT words which have the PALM vowel in a GA-style pronunciation (14 predicted occurrences),
- BATH words which have the TRAP vowel (3 occurrences),
- rhoticity (15 occurrences),
- t-voicing (12 occurrences),
- yod-dropping (1 occurrence),
- suffix vowels (1 occurrence),
- individual lexical items (3 occurrences).

This gave a total of 49 occurrences of the salient features, which were analysed in the reading of each speaker.

2.3. Speakers

We recorded 47 freshers at the Department of English and American Studies. Few of them had had any pronunciation training at school. They were recorded during their second pronunciation class at university, so their readings can be considered fairly representative of Bulgarian English pronunciation upon graduation from secondary school.

Three recordings were excluded from the analysis: two of them – because the students were aiming at an Australian accent, and one – because of bad quality of the recording. Thus, forty-four readings of the diagnostic passage were analysed in all. Thirty-four of the speakers were female, and ten were male.

2.4. Recording procedure

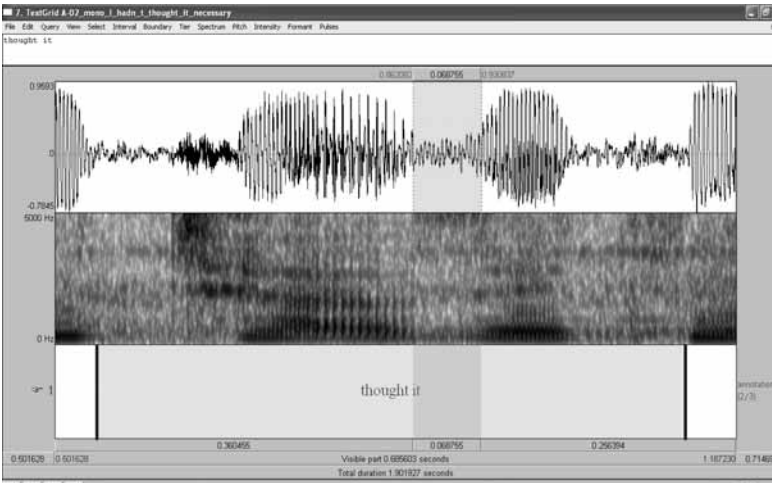
The recordings were made in the language lab in which pronunciation classes normally take place. Students recorded themselves simultaneously on their individual tape-recorders. The recordings were then transferred onto a master tape, digitized and stored on disk.

2.5. Analysis and discussion

The forty-four recordings were analysed both auditorily as well as with reference to the signal waveforms and the sound spectrograms in Praat (Boersma and Weenink 2012). The aim was to establish how consistent learners were in their use of the salient segmental features listed above. We listened to each recording as many times as necessary, until we felt certain that the feature in question had been identified correctly. Unfortunately, at times this turned into a rather daunting task, especially with respect to features such as rhoticity and t-voicing.

The realisations of intervocalic t varied a lot, from a voiceless plosive with normal closure duration to realisations with a very short closure stage, to voiced and/or tapped variants, fricative realisations, etc., some of which were particularly hard to classify. The screenshot in Fig. 1 shows a case of lenition and frication of the intervocalic /t/ in the phrase *thought it*.

Figure 1. The phrase *thought it* pronounced by speaker S2: the fricative realisation of the intervocalic t has been highlighted.

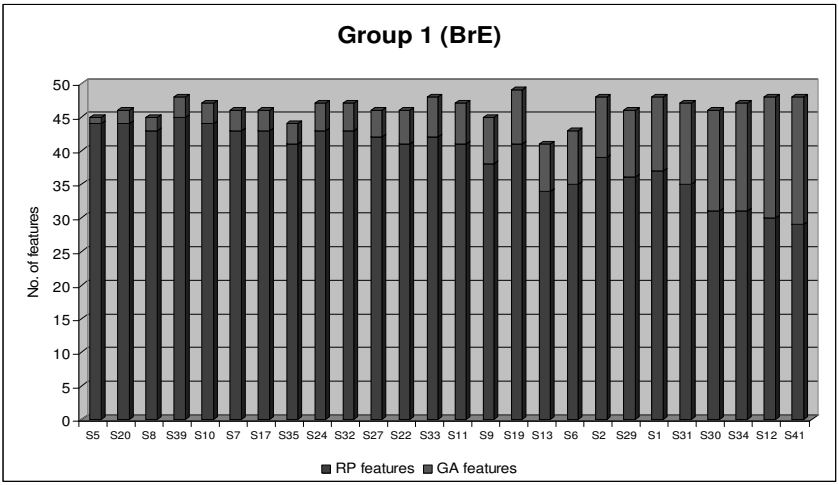


Some of the hardest cases to analyse involved rhoticity, especially those occurrences in which the auditory analysis was inconclusive and had to be supplemented by acoustic analysis. Ladefoged describes the English /r/ in *red* as being acoustically "usually marked by a decrease in the frequency of F3. Variations in the frequency of F3 indicate the degree of r-coloring: the lower the F3, the greater the degree of rhoticity" (2003: 149). Sometimes, however, such F3 lowering was hard to detect, even with the help of the Formant Tracker incorporated in Praat. Several doubtful cases of this kind were ultimately excluded from the final analysis.

In a few additional cases, incorrect stress placement accompanied by vowel reduction obscured the characteristic feature that was meant to be elicited: for example, the vowel in the second syllable of *accommodation*, which was designed to elicit an instance of the LOT vowel, was pronounced by some speakers without secondary stress and with the vowel reduced to schwa. These realisations were left out of the analysis for those particular speakers.

Ultimately, the pronunciation consistency judgments for many speakers were based on analyses of the occurrence not of 49, but of a smaller number of salient features. The students' recordings were sorted into groups according to the accent for which they expressed a preference.

Figure 2. Pronunciation consistency of speakers who aim at a British-style accent: speaker number is shown along the horizontal axis and the number of diagnostic features - along the vertical axis.



The largest group comprised those students who said that they wanted to sound British: 26 students in all (Fig. 2). For the most consistent of them – Speaker 5 (S5) – a total of 45 realisations were analysed, of which she had only one American characteristic (she pronounced the word *colleg*e with the BATH vowel). The least consistent speaker in this group – S41, on the other hand, out of a total of 48 occurrences had 19 Americanisms, most of them involving rhoticity and t-voicing. The average number of American characteristics for this group was 7.3, out of a total average of 46.3 features included in the analysis, or 15.8%.

The second group consisted of 7 students who wanted to sound American. As can be seen from Fig. 3, the last of them - S36 - used the biggest number of American characteristics – 26 out of a total of 47. Her most obvious inconsistency had to do with the LOT vowel which she pronounced rounded, as in RP. On average, the speakers in this group produced 20 GA features out of a total mean of 46.9; that is, the American characteristics in their speech constituted 42.6% - less than half of the total number of realizations under investigation. In other words, the speakers aiming at an American accent were further away from the reference accent of their choice than those who chose a British accent, the latter being "off the target" only 15.8% of the time on average.

Figure 3. Pronunciation consistency of speakers who aim at an American-style accent: speaker number is shown along the horizontal axis and the number of diagnostic features - along the vertical axis

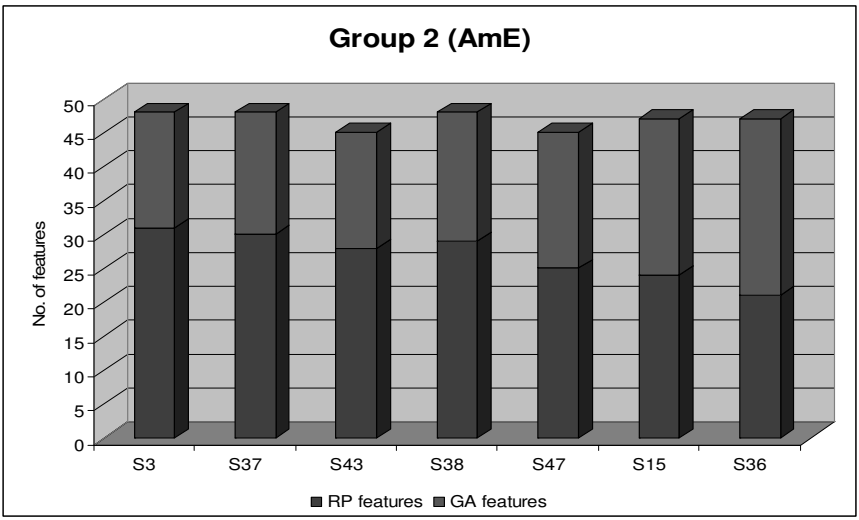


Figure 4. Pronunciation consistency of undecided speakers: speaker number is shown along the horizontal axis and the number of diagnostic features – along the vertical axis.

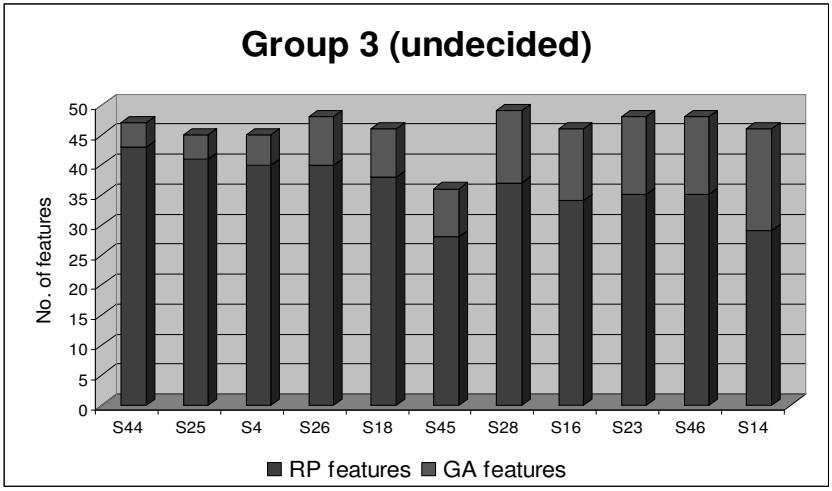
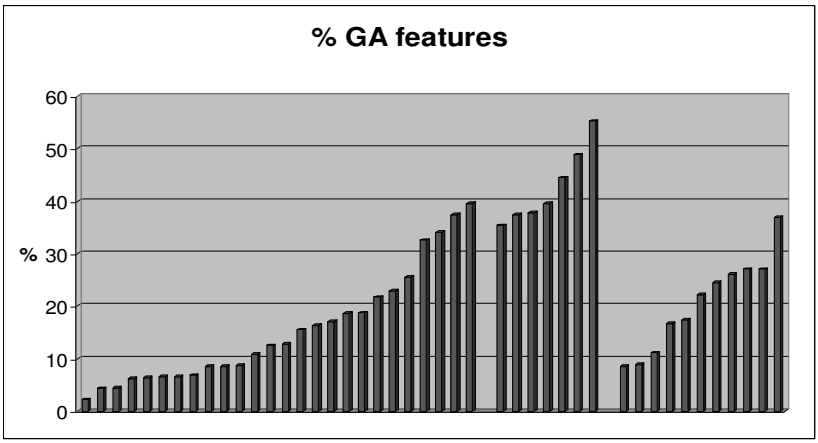


Figure 5. Percentage of American English pronunciation features for each speaker across the three groups; groups 1, 2 and 3 (from left to right) are separated by spaces.



The results for the last group of students – those who at the time of the study did not express a preference for any pronunciation model, are shown in Fig. 4. They generally resemble the results obtained for Group 1.

Finally, Figure 5 shows only the percentage of GA characteristics for each speaker in each of the three groups; the groups are separated by spaces.

Although group means were not tested for statistical significance (the groups were small and group sizes were diverse), the graphs in Figure 5 illustrate the differences between the three groups. In the first group, more than two-thirds (19 students) produced more than 80% of the salient features in accordance with the British-style pronunciation model of their choice. However, among the students in the second group who aimed at sounding American, only one speaker managed to do so with respect to the characteristics under investigation in more than 50% of the cases, and another one narrowly missed the 50% mark. On the whole, speakers from the first group – those who said that they wanted to sound British - were much more consistent in their attempt to adhere to the pronunciation model of their choice. It must be borne in mind, however, that group sizes were unequal, which may have had an influence on the results which we obtained.

With respect to the most salient characteristics which were elicited, rhoticity (or rather, the avoidance of rhotic pronunciation) was the characteristic feature which was most difficult for the first group of speakers to produce consistently, followed closely by t-voicing. For the group who had chosen GA as their reference accent, the biggest challenge was presented by features other than rhoticity and t-voicing, such as the consistent production of the typical American vowel qualities in words of the LOT and BATH lexical sets.

Finally, for all three groups, the pronunciation of individual words such as *clerk* and *leisure* turned out to be a formidable task. This, however, may be indicative of the unfortunate trend, still sometimes encountered in Bulgarian schools, to pay very little attention to pronunciation in general, let alone such “niceties” as the differences between British and American pronunciation.

3. Conclusion

Given the long tradition of teaching British-style pronunciation in Bulgaria, it may not be surprising that students imitate more consistently an RP type of pronunciation rather than a GA one. Nevertheless, pronunciation instruction at tertiary level should also be able to cater for

the sizeable minority of students who opt for an American pronunciation model. The problems with which we are faced when we teach native-speaker pronunciation in a non-native environment are numerous. While tertiary level students in English departments in Bulgaria and elsewhere continue to aim for native-like pronunciation, analyses like the one reported in this paper would hopefully add to our understanding of students' problems, as well as to the insights already provided to us by Contrastive Analysis, Error Analysis, etc. It is our belief that the problems posed by the requirement of consistency are no lesser than problems created by mother tongue interference. The teaching situation as far as pronunciation is concerned is extremely complex nowadays, due to the fact that we are living in a diverse multi-cultural environment in which many different Englishes are spoken, which makes it much more difficult for our students to focus on their target accent. However, a better knowledge about our students' special problems with respect to accent consistency will hopefully enable us to help them achieve their pronunciation goals.

References

Boersma, P. and D. Weenink. 2012. Praat: doing phonetics by computer [Computer program]. Version 5.3.17. Retrieved 12th June 2012 from http://www.praat.org/.

Collins, B. & I. Mees. 2008. *Practical Phonetics and Phonology. A resource book for students*. Second edition. London and New York: Routledge.

Cruttenden, A. 2008. *Gimson's Pronunciation of English*. Revised by A. Cruttenden. Second edition. London: Hodder Education.

Dimitrova, S. 2003. *English Pronunciation for Bulgarians*. Sofia:Vezni-4.

Dimitrova, S. & Ts. Chernogorova. 2012. English pronunciation models and tertiary-level students: a Bulgarian perspective. In *Exploring English Phonetics,* edited by B. Čubrović and T. Paunović, 199–215. Newcastle Upon Tyne: Cambridge Scholars Publishing.

Dziubalska-Kolaczyk, K. & J. Przedlacka (eds.). 2005. *English Pronunciation Models: A Changing Scene*. Bern: Peter Lang.

Kortmann, B. & C. Upton (eds.). 2008. *Varieties of English, vol. 1. The British Isles*. Berlin: Mouton de Gruyter.

Ladefoged, P. 2003. *Phonetic Data Analysis. An Introduction to Fieldwork and Instrumental Techniques*. Oxford: Blackwell Publishing.

Ladegaard, H. & I. Sachdev. 2006. "I Like the Americans… But I Certainly Don't Aim for an American Accent": Language Attitudes, Vitality and Foreign Language Learning in Denmark. *Journal of Multilingual and Multicultural Development* 27(2): 91–108.
Llurda, E. 2009. Attitudes towards English as an international language: the pervasiveness of native models among L2 users and teachers. In *English as an International Language: Perspectives and Pedagogical Issues*, edited by F. Sharifan, 119–134. Bristol: Multilingual Matters.
Roach, P. J., J. Setter & J. Esling (eds.). 2011. *Cambridge English Pronouncing Dictionary by Daniel Jones*. Cambridge: CUP.
Schneider, E. W. (ed.). 2008. *Varieties of English, vol. 2. The Americas and the Caribbean*. Berlin: Mouton de Gruyter.
Wells, J. C. 1982. *Accents of English*, 3 vols. Cambridge: CUP.
Wells, J. C. 1997. Whatever happened to RP? Retrieved 20th March 2012 from http://www.phon.ucl.ac.uk/home/wells/rphappened.htm.
Wells, J. C. 2008. *Longman Pronunciation Dictionary*. Third edition. Harlow: Pearson Longman.

Appendix 1

Diagnostic passage

I arrived in New South College on a Sunday afternoon. The porter at the lodge told me how to get to the central office block, where a clerk in the Accommodation Office gave me my keys. So I wandered about, looking for the pretty little cottage I had seen in the colour photograph in the prospectus. I hadn't thought it necessary to ask the clerk for directions. But it was getting dark and there was just nobody around. The beautiful blonde girl I had momentarily seen a minute ago had disappeared in the direction of the car park. Everybody seemed to have gone to spend their leisure time in the city. The dark green bushes on both sides of the path were beginning to look hostile, and I couldn't help thinking that I had got lost.

SLAVIC ENGLISH ACCENTS REVISITED: A CASE STUDY OF RUSSIAN SERBIAN-ENGLISH IN FILMS[1]

BILJANA ČUBROVIĆ

Outline

This chapter looks into the ways Serbian foreign accent in English is exploited in film industry, as represented by one of the internationally renowned actors of South Slavic language background. Carefully selected audio recordings are analyzed from a segmental viewpoint, with the help of acoustic phonetics tools, and also from an auditory perspective, where necessary. The main aim of this chapter is to explore how a non-native accent as Serbian is twisted for this purpose, and which phonetic features, if any, are most suited to achieve an atmosphere of a strong Russian accent in English.

1. Introduction

In an attempt to find the most suitable actor for a role, movie crews seem to rely on resources most readily at hand at a specific moment in time. Suppose you are based in Hollywood and you need an actor to play a Russian mobster as soon as possible. Which type of accent would be considered heavy enough to perform this task, provided you do not have a Russian native speaker at hand? The best guess is an accent which can be classified broadly as a Slavic one. For the purpose of proving this method right or wrong, I have chosen to study the English spoken by one of the most accomplished actors of South Slavic background, Rade Šerbedžija, in several recent American and British film productions. His foreign accent may be labelled as strong, but it seems suitable for the roles impersonating

[1] This chapter is part of the project no. 178019 supported by the Serbian Ministry of Education, Science and Technological Development.

Russian villains. The characters played by Šerbedžija are indeed very often of Russian origin and the analysis to follow will be based exclusively on such roles.

For one of the films studied here, namely *Mission Impossible 2*, there is information available that Šerbedžija had received some instruction from a renowned dialect coach Victoria Mielewska based in Australia, as stated on her personal website[2]. The implementation of professional dialectal guidance for other movies analyzed in this chapter is vague, and no details about it are available to the author of the present chapter.

1.1. Critical Period Hypothesis

It is a commonplace in linguistics to say that after an individual has reached a certain age, they will no longer be capable of learning a foreign language to the extent that their foreign accent is completely annihilated in the language acquisition process. Linguists disagree about the exact age until which a full mastery of a language is possible, but most mention puberty as a delimitating period. This is often referred to as the Critical Period Hypothesis, first proposed for L1 acquisition by Lenneberg (1967), where he claims that the critical period starts from about age 2 and ends with puberty. The Critical Period is also proposed for L2 acquisition, but various other factors alongside with age must be taken into account - motivation, cultural empathy, desire to sound like a native speaker, and type and amount of L2 output (Major 2001: 9). It is a fact of life that certain language communities keep tightly together when living in an L2 environment, and this may hugely affect the level of mastery of the foreign language in question.

When conversing with native speakers of an L2, children usually acquire the language without problems, but for most adults this is an extremely difficult task, often next to impossible. Also, automatic acquisition from mere exposure to a given language seems to disappear after puberty, and foreign languages have to be taught and learned through a conscious and laboured effort. According to Lenneberg (1967: 176) foreign accents cannot be overcome easily after puberty, and failure rather than success is associated with second language acquisition (SLA).

[2] See http://www.creativevoice.com.au/Creative_Voice/Creative_Voice_Victoria_Mielewska_Home.html.

Selinker (1972) claims that a mere 5% of learners attain absolute success when learning an L2, a fact that is not at all encouraging.

Speaking from a listener's perspective, listeners make judgements (consciously or unconsciously) about whether someone is a native or non-native speaker of their language (Major 2001: 19). Longer stretches of speech are much better for such evaluations because under such circumstances it is impossible to avoid various problematic segmental and suprasegmental phenomena. In highly controlled speech one can avoid speech sounds that they find troublesome and avoid being labelled as a non-native speaker. The problem with such judgements is that they are hardly ever objective (Markham 1997: 85). What further complicates the matter is the fact that dialectal differences are often heard as non-native, and foreign influences are perceived as dialectal. As Major (2001: 12) puts it "[T]he notion pass for native is not a simple matter [...]. Does it mean fool some of the people some of the time, all of the people some of the time, or all of the people all the time?". However, listeners are sensitive to subtle differences in phonetic structure contributing to foreign accent (Magen 1998: 381). And I will now proceed to describe the phonology of Russian English, also drawing parallels to Serbian English, wherever necessary.

1.2. Interlanguage: Russian vs. Serbian in English

According to Swan and Smith (2001) Russian learners of English encounter numerous problems when pronouncing English. They offer the account of problems Russian native speakers have when learning English, more specifically a standard British variety of English. The two major features which distinguish the Russian phonological inventory from English are the lack of short-long vowel differentiation and the absence of diphthongs, as well as differences in English rhythm and stress patterns (Swan and Smith 2001: 145). Namely, the English vowel system is much more intricate than the one characteristic of Russian. Vowels like /ɜː ɑː æ ɔː/ do not have equivalents in the Russian vocalic inventory. The most common transphonemic paths operating between English and Russian are:

(1) Eng. /ɜː/ → Russ. /э o/ (The Russian /э/ is similar to Eng. /e/)
(2) Eng. /ɑː/ → Russ. /a/
(3) Eng. /æ/ → Russ. /e/
(4) Eng. /ɔː/ → Russ. /o oʊ/

The second elements of English diphthongs tend to be overpronounced by Russian native speakers. The difference between long and short vowels of English is difficult to master, and long vowels are often insufficiently "tense" (Swan and Smith 2001: 146).

The English consonants /θ ð ŋ w h dʒ/ are often mispronounced by Russian native speakers due to their absence in the Russian consonantal system. /θ ð ŋ w/ prove very difficult for a Russian native speaker, similar to native speakers of Serbian. The dental fricatives are often replaced by /s z/, unlike Serbian where native speakers mispronounce them as /t d/ in Serbian-coloured English speech. The English velar nasal /ŋ/ is realized as /g/ or even /n/, which is not the case in Serbian English where only /g/ is often added in medial and final positions. The failure to produce English words that contain /w/ results in the same mispronunciation used by Russian and Serbian speakers of English. Furthermore, English /t d n l/ are pronounced as dental by native speakers of Russian due to which native speakers of English may characterize them as foreign. Final devoicing is common in Russian, and this phonetic habit is transferred into English. Russian speakers of English would often say *lap* instead of *lab*, for instance. Similar to Serbian speakers (Čubrović 2011; Čubrović 2012), Russian speakers of English do not insist on aspiration of voiceless plosives in English. One of the most striking features of Russian-coloured English is the tendency of a Russian speaker to palatalize most English consonants before front vowels like /iː ɪ e eɪ ɪə/ (Swan and Smith 2001: 147). Palatalization of English consonants by Russian learners is a feature that is not induced in Serbian-coloured English speech, and this is where Russian native speakers deviate from Serbian EFL learners. Furthermore, velarized [ɬ] is used instead of the palatalized one in most phonetic environments in English as produced by a Russian native speaker. This interlanguage characteristic is not found in Serbian EFL speech. Finally, the English /tʃ/ is often palatalized in Russian English, which is not the case in Serbian. The account provided by Swan and Smith (2001) is very exhaustive when it comes to the differences between English and Russian phonological inventories, but they do not claim that every Russian EFL learner will accommodate all of them. I have chosen only the most important differences, which may prove useful for the analysis of the corpus that follows. Suprasegmental differences are not taken into account due to the nature of the corpus that is mostly made up of isolated words and very short utterances.

In his instruction manual for 24 accents and dialects commonly used by English-speaking actors, Meier (2002: 133) points out that there is no such thing as "the Russian accent, only a Russian accent". Russian is one

of the foreign languages Meier (2002) includes in his manual alongside with Afrikaans, French, German, Italian, Spanish and Yiddish. He also makes recommendations about how to pass as a person from different areas in the United Kingdom (Cockney, Hampshire, Ireland, Liverpool, Northern Ireland, Scotland, Wales or Yorkshire, or be a speaker of RP), or the United States (American South, Deep South, Downeast New England, New York, South Boston, or speak with a neutral GA accent).

When listing what he refers to as the most common pronunciation problems an actor might encounter when challenged to speak with a Russian accent, Meier stresses that it is highly unlikely that one person will exhibit all the problems he elaborates on. In other words, there will always be a divergence from one speaker to another. According to Meier (2002: 133–135), among what he calls 'signature sounds' are the English /ɪ/ in which case the Russian closest approximation /i/ is often used. The Russian vowel /i/ is often slightly longer than the English one. Similarly, the English /ʊ/ is often substituted by a closer, more rounded /u/. The unrounded back English monophthong /ʌ/ gets replaced by the Russian rounded back vowel /ɒ/, as in *country, among, monk*. This substitution may also be caused by the pronunciation based on the spelling of these English words. As there is no /æ/ in the Russian vowel inventory, this English vocalic sound is substituted by /ɑ/ or /ɛ/, as in *program, bad, mankind.*

English diphthongs may pose problems to Russian native speakers, where /eɪ/ changes into a long e-vowel, e.g. *danger, label, wave*. In the same vein, the diphthong /əʊ/ may sound more like a monophthong which is similar to /ɔ/, e.g. *mobile, broken, hopeless*.

When it comes to consonants, the most striking difference between Russian and English is associated with the 'r' sounds in the two languages. The Russian trilled /r/ is an unmistakable indicator of a strong foreign accent. There is also a tendency to mispronounce the English /w/ as /v/ in Russian-coloured speech. Meier (2002: 135) emphasizes that what he calls the "light [l]" in English is a most difficult sound for Russian speakers which often resembles a dark [ɬ] that is common in Russian. The English th-sounds are no exception. Russian EFL speakers often confuse them with their own /t/ or /d/ sounds that are dental. Another empty slot in the Russian consonantal inventory is the glottal fricative /h/, which is often substituted by the Russian palatal fricative [ç] or even [x].

Lastly, Meier's extensive experience with actors shows that pronouncing -ing forms (and other English words containing this sequence of sounds, or -ong) with the velar plosive /g/ is a sure indicator of a Russian foreign accent in English, e.g. *king, thinking, song*. Intrusive [j]

can further contribute to an actor sounding more like a Russian speaker. Some of the phonetic environment where an actor might find it useful to insert a palatal are: after /m/ and /n/, and before [i], [ɪ] and sometimes [e], e.g. *meter, enable, beneath, meeting, needful, meaning, never, main, street* (Meier 2002: 135).

The most notable additional features that a Russian native speaker will incorporate into their English are: devoicing of voiced consonants (plosives and fricatives in word-final positions, dentality of /t d n/, and lack of aspiration of /p t k/ (Meier 2002). Various other suprasegmental differences are likely to manifest in what we perceive as Russian-coloured speech, but we will set them aside.

3. A note on the corpus

The analysis presented in this chapter is based on four films featuring the actor Rade Šerbedžija, in his minor roles, where he is asked to imitate an accent resembling a Russian one in his English speech. All four productions are relatively recent and they are listed in Table 1:

Table 1. Data on the corpus

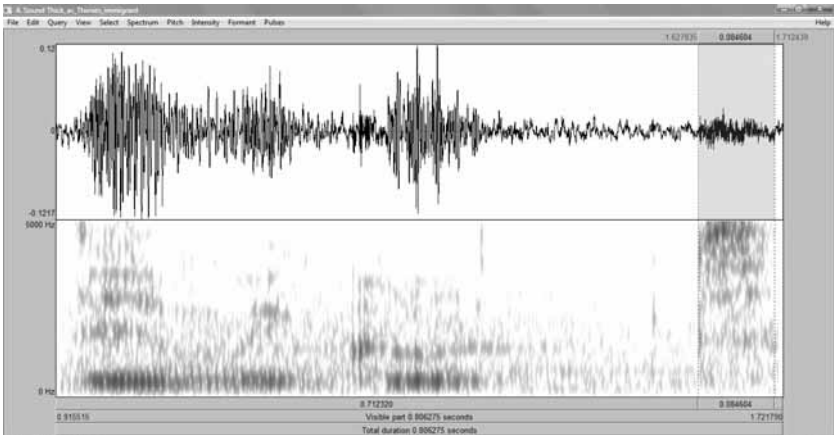
Marked as (the number given below) in the text	**Year**	**Title**	**Role**	**Description**
(1)	1997	*The Saint*	Ivan Tretiak	A former Communist party boss and a billionaire oil and gas oligarch
(2)	2000	*Mission Impossible 2*	Dr. Vladimir Nekhorvich	A Russian bio-chemical expert
(3)	2000	*Snatch*	Boris Yurinov	An ex-KGB agent, aka "The Blade"
(4)	2009	*Thick as Thieves*	Victor (aka Nicky)	A Russian mobster

4. Corpus analysis

4.1. Consonants

The auditory analysis of the corpus allows categorization of the data leading to a foreign accent into several groups. The most notable characteristics of the actor's speech that allude to an existence of a strong foreign accent concern consonants. Firstly, as expected the /r/ sounds are pronounced as alveolar trills or taps most of the time, in words like *freeze* (1), *provide* (1), *ridiculed* (1), ***r**evered* (1), *runs* (1), *entrust* (2), *blood pressure* (2), *requires* (2), *reliability* (3), *drop* (3), and *Russia* (3). It is also worth mentioning that in many sequences in which Rade Šerbedžija speaks out, there is some kind of underlying auditory effect (music, noise, etc.) and acoustic analysis is hard to perform due to a high level of noise. In such instances, we need to rely on what we hear. If the sound in question is pronounced as a strong trill, there is not doubt about its manner of articulation. However, taps are sometimes hard to spot due to the background noise.

Another feature of Šerbedžija's Russian is a notable full release of final plosives in *out* (1), *quiet* (1), *viewpoint* (1), *that* (2), *fact* (4), *left* (4), *immigrant* (4), and *hope* (4). In most instances, the fully released plosive is a voiceless alveolar /t/, and its typical full release is realized as part of an added syllable. This syllable addition creates an impression of a foreign accent, which is probably an effect the actor wants to achieve. The shaded area in the wave form clearly shows the release of the plosive /t/ in the English word *immigrant*:



In connection to fully released stops that native speakers of English realize, this feature of plosives is not at all uncommon. Full releases are typical of enunciated or very careful speech. However, in casual connected speech, final plosives usually have an inaudible release.

Next, a noticeable substitution of the English alveolar /t d/ in typical /tr dr/ consonant clusters is evident in the corpus. In the lexemes *treasury* (1), *traitor* (1), *entrust* (2), *drop* (3), the initial consonant is pronounced as dental, which is a feature typical of both Russian and Serbian interlanguage. However, in *countrymen* (1), the consonant in question resembles the English alveolar /t/, but it also very close to a common Serbian EFL learner's substitution of plosive for a similar affricate.

The English velar nasal is invariably followed by a velar plosive realized in the following words from the corpus: *doing* (1)*, spinning* (1), *selling* (1), *long* (1), *nothing* (3) and *bring* (4). Such an addition is not characteristic of English, at least not in final positions. However, both Russian and Serbian EFL learners commit to it, which again alludes to a strong foreign accent in English.

Complete devoicing of the final voiced alveolar fricative occurs in *requires* (2)*, days* (3)*, bookie's* (3). In those instances, the final consonant is lengthened which makes it even more foreign-sounding. This change is also mentioned by Swan & Smith (2001: 147), Meier (2001: 136) and Meier (2011: 306) in their detailed accounts of a Russian accent in English. Meier lists it under what he refers to as 'additional features' which create a strong impression of a foreign accent, and he also points out that the final consonant devoicing affects the shortening of the preceding vowel, which is "perhaps the 'foreignness' that our ear detects most easily" (Meier 2011: 306). Serbian EFL speakers at a very low level of English proficiency make the same mistake, but this is mostly due to the influence of orthography.

4.2. Vowels

The most noticeable characteristic of Šerbedžija's Russian accent is the lack of vowel reduction in numerous English words, e.g. *Oxford* (1), *lunatic* (1), *provide* (1), *Atlanta* (2), *possession* (4). A full vowel is used instead of a schwa in the instances where the vowel is underlined. This may be an influence of the spelling on pronunciation of such words.

Moreover, Šerbedžija applies vowel shortening in words like *easy* (1) or *bookie's* (3). On the other hand, a vowel lengthening is perceived in *detectors* (1), and *Russia* (1). Vowel shortening could be accounted for by Russian phonology, as pointed out by Swan & Smith (2001: 146). Vowel

lengthening might be an idiosyncratic feature of the actor's interlanguage phonology.

5. Final remarks

In Pedersen's study (2007) of linguistic credibility of Mel Gibson's Scottish accent in the film *Braveheart*, he aptly points out that there seems to be emerging a trend to imitate foreign accents in Hollywood. However, how sincere Hollywood is in its intentions and how big an effort is involved when it comes to promoting language variation in movies, is yet unclear. The efforts of this sort are commendable, and there are many benefits that English-speaking audience could gain from a mere exposition to other language idioms. When discussing Mel Gibson's imitation of Scottish, especially in relation to the Scottish Vowel Length Rule, Pedersen (2007) points out that the actor has obviously put some effort into sounding like a Scotsman, but still some inconsistencies are evident when it comes to deviations from Gibson's native GA dialect. Gibson therefore tries hard, but he would not be mistaken for a native speaker of a Scottish dialect. However, it seems that Gibson's accent is "adequate for the international market" (Pedersen 2007: 89). What "adequate" is in this sense of the word is not very transparent. Pedersen probably refers to movie audiences not prone to spotting every single "mistake" an actor makes when imitating a Scottish (or any other) accent. The overall impression should be that of sounding Scottish occasionally, and that will suffice. Furthermore, films are not made for linguists, but for laymen, and they are not to be scrutinized, but watched and enjoyed.

As for Šerbedžija's Russian idiolect, the author of this chapter agrees with Pedersen's views on Gibson's Scottish in that Šerbedžija's Russian accent offers some local Slavic colour, though he fails to imitate most easily noticeable features of Russian English. One characteristic that Serbian and Russian learners of English share is the trilling quality of /r/'s, and Šerbedžija utilizes it to a large extent, though he is not always consistent. One striking feature which Serbian and Russian speakers do not share when pronouncing English is the palatalization that is found in certain phonetic enviroments. Šerbedžija does not attempt to utilize this feature, which would undoubtedly lead to the impression of a strong Russian foreign accent in English. On a tentative scale of foreignness of features that would lead to a heavy Russian accent, palatalization of several consonants explained in section 1.2. would be rated high. The actor, however, does not show any inclination towards its usage, but he stills sounds like a Russian character, most of the time. This seems a

satisfactory result for the movie audience, but from a linguistic viewpoint, his dialect coach could have contributed more. It seems that gaining a reputation for playing Russian characters made his dialect coach more lenient.

References

Čubrović, B. 2011. Voice Onset Time in Serbian and Serbian English. *English Language Overseas Perspectives and Enquiries* VIII, Spring: 9–18.

Čubrović, B. 2012. Evolution of a Phonetic Phenomenon: The Case of Voice Onset Time in Serbian Intermediate EFL Learners. In *Exploring English Phonetics*, edited by T. Paunović & B. Čubrović, 47–55. Newcastle upon Tyne: Cambridge Scholars Publishing.

Lenneberg, E. 1967. *Biological Foundations of Language*. New York: Wiley.

Magen, H. S. 1998. The Perception of Foreign-Accented Speech. *Journal of Phonetics* 26: 381–400.

Major, R. C. 2001. *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ/London: Lawrence Erlbaum Associates.

Markham, D. 1997. *Phonetic Imitation, Accent and the Learner*. Lund: Lund UP.

Meier, P. 2002. *Accents and Dialects for Stage and Screen: Some of the Accents and Dialects Most Commonly Used by English-Speaking Actors in Film and in the Theatre*. Third edition. McLouth, KS: Paul Meier Dialect Services.

Meier, P. 2011. *Accents & Dialects for Stage and Screen: An Instruction Manual for 24 Accents and Dialects Commonly Used by English-speaking Actors*. Lawrence, KS: Paul Meier Dialect Services.

Selinker, L. 1972. Interlanguage. *IRAL* 10, 2: 209–231.

Swan, M. & B. Smith. 2001. *Learner English: A teacher's guide to interference and other problems*. Cambridge: CUP.

PHONETIC SIMILARITY IN TRANSLITERATED ENGLISH TRADEMARKS – A PRELIMINARY STUDY IN JAPANESE

ISAO UEDA

Outline

The present study attempts to discuss phonetic and phonological problems applicants of new foreign trademarks encounter in legal decisions. First, a brief explanation of the trademark and application procedure is provided. Next, we look at examples which sound (and look) completely different to the foreign applicants but which were in fact judged to be phonetically similar and turned down by Japanese examiners. It is suggested that the sole reason for this apparent discrepancy is caused by the Japanese trademark law which demands newly proposed foreign trademarks to be transliterated into Japanese orthography. This process can result in the original form, being distorted phonetically. We then observe some examples of trademarks and argue that the similarity decision is affected by certain factors, such as segmental contents, the different position of the syllable in a word and the total length of the trademark. Finally, based on an example of apparent inconsistent judgment, a suggestion is made to improve the current situation.

1. Introduction

We have all heard of the word “trademark”, while many of us are unfamiliar with what it means exactly. Let us start the discussion by looking at how “trademark” is defined. *Longman Dictionary of Contemporary English* (1987) gives the following definition.

(1) Definition of "trademark"

> A special name, sign, word, etc. which is marked on a product to show that it is made by a particular producer, and which may legally only be used by that producer.

Even if we did not know the exact meaning of the word itself, there are a whole range of trademarked products which we encounter in our daily life. For example, McDonald's (food), Minute Maid (drinks), Tiffany (jewelry), SONY (electrical appliances), Toyota (automobile), Apple (computer) and Panasonic (electrical appliances), to mention but a few. Trademarks are also labeled on services such as Diners Club (credit card), Federal Express (logistics), Holiday Inn (hotel), etc. Even a public institution such as Osaka University may be registered as a trademark, so that they can be prevented from being used in an undesirable way. Here, what is important is that a trademark is *legally* protected: other producers, an individual or an organization, are not allowed to use the (phonetically) same name for a new product if it has already been registered as a trademark. In the examples cited above, Apple used to be a music record company established by the former Beatles members. To use the name, an American computer company offered an enormous amount of money; they "bought a trademark". Panasonic, which started in Japan as "National", when they found their way into the American market, realized that there had already been a company registered as "National" in the United States, so they gave up their original name for "Panasonic". It is somewhat ironical that the same company recently abandoned their original name even in Japan and started using an integrated name "Panasonic". Hence, we cannot use a phonetically same name, label, service, or whatever, if it has been registered as a trademark.

Then, what happens to phonetically "similar" names? Trademark laws are different across countries, but Dinnsen (1987) explains the situation in the United States as follows.

> (2) Under federal and states statutes, the name of a product or service is protected against infringement by others where the similarities are such that confusion between names (and thus products) could result. Thus, a proposed new product name may illegally infringe on an established name if it is judged to be too similar in sound (Dinnsen 1987).

Here a question comes up almost immediately. If, as the rule goes, a newly proposed name infringes on an existing name by having been judged as similar, how is that judgment made? How similar is similar and how different is different? If possible confusion between an established name and a proposed name is a source of infringement, how do we decide that the two names are confusable with each other? What are the criteria? Dinnsen (1987) adds that the law sets down no explicit or objective criteria for assessing phonetic similarity and likelihood of confusion. It is easily and logically expected from this that the decision of similarity and confusability may be different amongexaminers or judges, and if that is true, it is not a desirable situation because it results in lack of fairness under the law.

In Japan this putatively possible inconsistency is, unfortunately, observed and recorded at least in some lawsuits, which resulted in unfair judgments. Given these circumstances, the purpose of the present study is as follows. First, as a preliminary attempt, we examine how phonetic similarity is determined on the basis of phonetics and phonology. Here, we will focus only on English trademarks applied for registration in Japan. This is because many of the puzzling and unfair court decisions have been witnessed in the application of such foreign trademarks. Next, in the process of searching for the problems concerning foreign trademarks, a "gap" in awareness between applicants from abroad and the Japanese authorities is brought into relief. It is the difference in phonological structures between Japanese and English (and other Western languages) which causes the problem. We then turn to sets or pairs of cases in which apparently discrepant judgments have been made by the court, which is followed by a discussion of why it is a problem from phonetic and phonological viewpoints. Finally, we conclude that the decision of phonetic similarity is affected by phonetic, phonological and cognitive variables, with a brief suggestion for a better court decision. Overall, the study is expected to serve as the first step to reach the ultimate, far-off goal of establishing objective norms to judge phonetic similarity.

2. Phonetic similarity and source of confusion

As mentioned above, there is a difference in the system of trademark laws across countries.

In the United States or the United Kingdom, potential confusion is determined by the potentials of language to confuse and by surveys showing the actual confusion of consumers. In Japan, on the other hand,

the likelihood of confusion is determined by how the law is read, not by linguistic analysis or market survey research (Ogawara 1999, quoted in Shuy 2002). Then, how is the decision of phonetic similarity made in Japan? Under the present trademark law, three criteria are established to evaluate proposed trademarks. These are "sound," "appearance," and concept." The final determination is made by examining all of these three aspects of the proposed trademark. In this study, we discuss only "sound", which is relevant to our focus of attention, but we should not leave the other two out of consideration.

Let us start our discussion of phonetic similarity of foreign trademarks. In the part touching on "sound" or phonetic similarity, the trademark law goes as follows.

> (3) The decision of phonetic similarity should be made ... by overall perceptual impression when the two trademarks are pronounced. (Trademark Law 4.1.11)

It is easy to imagine that "overall perceptual impression" can and may vary across individual listeners. Thus, in the first place, the law specifies no clear, objective criteria by which phonetic similarity is to be judged.

Another point to be made is that when a trademark application is filed for a mark from a language such as English, the examiner first transliterates the reading of the mark into Japanese. Here comes the problem caused by the difference between English and Japanese phonological systems. In some cases, two trademarks which sound and look totally different to native speakers of the original language may be judged as phonetically similar by Japan Patent Office examiners. We will look at some puzzling examples cited in Hall (2004) below, but before we do, let us take a brief look at the procedure of examining applied trademarks.

When a new trademark application is submitted, it undergoes the first examination. If it is granted permission, it is registered as a new trademark. However, if it is rejected, then the applicant either gives up or asks for reexamination, after which it is either accepted or rejected again. If it is accepted, it is entitled to be used as a new trademark, and if rejected the applicant is given no further chances to "win the game".

Now let us observe three examples from Hall (2004). They were all applied for from abroad but all were considered to have phonetically similar "predecessors", and they failed the first examination. In (4) the

applied trademarks are juxtaposed with the already existing trademarks judged to be phonetically similar.

(4) Newly proposed trademarks Already established trademarks

1	RUBRA	v.	LABRA
2	PLEXTOR	v.	PRECOSTER
3	PIGEON	v.	VISON

To Hall's surprise, all of the newly proposed trademarks above were rejected in the first examination. It is a great surprise to Hall because none of the three pairs seems phonetically similar from the viewpoint of a native speaker of English. In the next session, we discuss why these seemingly "phonetically dissimilar" pairs caused a confusion problem.

3. Source of confusion

Why is a trademark judged to be phonetically similar and thus confusable by the Japanese judges when, to native speakers of English, it seems completely different from its existing counterpart? One significant reason is this; when a trademark application is filed for a mark in a language such as English, the examiner first transliterates the reading of the mark into Japanese. In this process, it is rephonologized and nativized into Japanese. Then, it is probable that the difference in the phonological structure between English and Japanese affects the reading to some extent. Let us see three relevant aspects of difference in phonological systems of Japanese and English.

First, segmental inventories of the two languages are quite different. Here is the segmental inventory of Japanese.

(5) Segmental inventory of Japanese

1 Vowels

i, e, a, o, ɯ

2 Consonants

Plosives: p, b, t, d, k, g
Fricatives: ɸ, s, z, ʃ, ʒ, ç, h
Affricates: ʦ, ʣ, ʧ, ʤ
Nasals: m, n, ɲ, ŋ, ɴ
Liquid: ɾ
Approximants: j, ɰ

The number of vowels is considerably smaller than that in English. That is, there are only five vowels in Japanese. This is expected to cause a problem when foreign words are adopted from a language like English with a relatively large inventory of vowels. Many English vowels in original words are neutralized in their adopted counterparts. This is especially notable in low-to mid-low region in the vowel space. Thus, "pat" and "putt" have been adopted to be homonyms, as with "collar" and "color". The same holds true for consonants. Although consonants appear rich in number, only p, b, t, d, k, g, s, ʃ, h, ʤ, m, n, ɾ, j, ɯ can be considered independent phonemes and the rest are allophonic varieties of these phonemes; phonemes are relatively few. Therefore, no distinction is made between English [l] and [r], [s] and [θ], nor between [b] and [v]: "bolt" and "volt" and "lock" and "rock" are homonyms in their adopted forms. The point here is that some English vowels and consonants undergo merger in the adopting process, resulting in a possible cause of phonetic similarity.

Second, the syllable structure of Japanese is considerably different from that of English. While English permits a wide range of syllable types, Japanese basically allows (C)V, and (C)VC where the coda has to be a nasal. To compromise with complex syllable types of English, vowels are usually added in loan words. Thus, "McDonald" pronounced with three syllables in English is pronounced with six syllables, something like [ma.ku.do.na.ru.do] in Japanese. (The final {'s} is not pronounced in Japanese.) It should also be pointed out that the syllable structure is closely related to the orthographic system of Japanese. Japanese has three types of letters and characters. They are *Kanji*, ideogramic Chinese characters, *Romaji*, phonogramic Roman alphabet, and *Kana*, which are syllable-based letters. There are two types of *Kana*, which are *Hiragara* used to denote traditional referents and *Katakana* used to denote recently adopted referents mostly of foreign origin. Below, some examples are given.

(6) Examples of *Kana* letters
1 *Hiragana*
あ[a], と[to], び[bi], め[me], や[ja]
2 *Katakana*
ア[a], ト[to], ビ[bi], メ[me], ヤ[ja]

It is into this *Katakana* that applicant foreign trademarks are transliterated. They can undergo a "metamorphosis", segmentally and/or prosodically. Getting back to the case in (3), we can observe what

happened to the applied trademarks. In the first example, the proposed "RUBRA" was transliterated as 「ルブラ」which is pronounced as [ɾɯbɯɾa]. On the other hand, the existing trademark "LABRA" is pronounced as 「ラブラ」[ɾabɯɾa]. The difference is reduced only to the difference in a single vowel. The second example, "PLEXTOR" is transliterated as 「プレクスター」[pɯɾekɯsɯtaː] and "PRECOSTER" as「プレコスター」[pɯɾekosɯtaː]. Again, it is the difference of one vowel. By the same token, "PIGEON" in the third example is transliterated as 「ピジョン」[piʤoɴ], while "VISION" as 「ビジョン」[biʤoɴ], this time different by one consonant. Thus, what appears to be a great difference in English may turn out to be a small one in Japanese, which can or may be judged to be phonetically similar when transliterated.

To summarize, an applied English trademark may become phonetically similar to another already existing trademark in the process of rephonologization or nativization because of the difference in the phonological structure and by the trademark law practice which forces the mark to be transliterated in different orthography reflecting the syllable structure of Japanese. Therefore, it can safely be said that under the present trademark law, at least some applicants from abroad suffer an unexpected disadvantage.

4. Segmental factors affecting the confusability

In this section, we look at more examples to see phonetic and phonological factors concerning segments that might affect an examiner's decision, which in turn might lead to dubious judgments. The examples are pairs of trademarks, either accepted or rejected, all taken from Enuenu-kikaku (2004). This database covers 7510 decisions made in the 57 years after World War II through December 2004. Since the number of the cases is very large, I shall only focus on the phonological contents of the segments consisting of the onset and nucleus of a syllable of the trademark. More specifically, I selected cases in which two trademarks looked similar as a result of one having an onset and the other lacking in the corresponding syllable. Some examples are given below.

(7)

amoɾe	v	kamoɾe
imeːna	v	nimeːna
ʃiseaːta	v	ʃiseɾaːta
ʤimɯɾeo	v	ʤimɯɾejo

In each pair the words appear phonetically similar to each other, but some were rejected because they were judged to be similar, while others were accepted because they were judged to be phonetically distinct, despite the apparent similarity. We examine why in one case the similarity was judgedso, while in another case it was not, referring to phonological contents of segments as well as prosodic factors.

First, let us consider the nature of vowels by comparing the following two pairs.

(8) takɯtasɯ v akɯtasɯ
tekɯniːɾɯ v ekɯniːɾɯ

In (8), the first pair was decided to be phonetically similar, while the second was not. The only difference is that the first pair has [a] but the second has [e]. This difference attempts us to think that the nature of vowels exerts influence on the decision. In fact, Sakata (2008) inquires into how the decision of similarity differs across vowels based on a large number of cases. She calculateed how decision of similarity varied among vowels indicating the following percentage for each vowel.

(9) Percentage of similarity (Sakata 2008)
/ɯ/ : 52.9%
/a/ : 50.8%
/e/ : 46.6%
/o/ : 35.5%
/i/ : 22.7%

Out of the five vowels, /ɯ/ should be left out because of its unique status. It is the underspecified, phonological vowel in Japanese (Tabata 1993). It is the default vowel which is added in foreign word adoption. Among the other vowels, a syllable which has /a/ sounds most “similar” to the examiners’ ears and a syllable which has /i/ sounds least similar, with /e/ and /o/ falling in between. This order corresponds with the sonority hierarchy, in which low vowels are the most sonorous and high vowels are the least sonorous. The difference in percentage in (9) suggests that the more sonorous the vowel is, the more similar the syllable sounds. Thus, the first pair in (8) with the most sonorous vowel /a/ in the first syllable sounds similar, while the second pair with /e/ in exactly the same place sounds different. Among the phonological contents of segments, the sonority of the nucleus vowel is one factor which may affect the similarity

judgment. This is somewhat contrary to the contention in Bladon and Lindblom (1981), Hood and Poole (1980), and Pickett (1957) who claim in some one way or other that acoustic differences of vowels are robust. Presumably, the more sonorous the nucleus vowel is, the less prominent the onset consonant in perception.

Let us briefly touch on consonants in the onset position. The onset is combined with the nucleus to compose a syllable. (Of course, the syllable either has or does not have an onset.)

Dinnsen (1987) claims that if the onset is an obstruent, it is likely to be confused with other obstruents, and that if it is a sonorant, then it is confusable with other sonorants; obstruents are not confusable with sonorants. However, he does not discuss the confusability of a syllable with or without an onset. There are two possibilities: a syllable with an obstruent onset is more likely to be confused with its counterpart without an onset; or a syllable with a sonorant onset is more likely to be confused with its counterpart without an onset. So far no study, in relation to trademarks, has been conducted to see which is the case. In our database cited above, no relevant examples have been found. However, we can predict that the latter should be the case, appealing again to the sonority hierarchy. A syllable with an obstruent onset is claimed to be more "eligible" than a syllable with a sonorant onset. In the transition from the onset to the nucleus, the sonority rises more radically in the former than it does in the latter. As the difference in sonority between the onset and nucleus becomes larger, the syllable becomes more eligible; it is more salient in perception. Therefore, we can predict that if the onset is an obstruent, the syllable is less likely to be confused with its counterpart without the onset. If the onset is a sonorant, it is more confusable with the syllable without the onset. We certainly need further inquiry to attest to the prediction and this particular problem related to the segmental contents of onset should be left open until we have acquired a reasonable number of examples.

5. Other prosodic and functional factors

The judgment seems to be affected not only by segmental factors as discussed above but also by prosodic or functional factors. The first of these is the position of the syllable in a word. Let us observe the following pairs.

(10) mijai v miai
jamɯko v amɯko

In both pairs the only difference is [ja] and [a]. However, the first pair was judged to be similar, and the second as different. This suggests that the difference in the first syllable is more prone to be judged to be different. This is because the first syllable in a word is perceptually most prominent and thus the difference is more noticeable than in other positions in a word.

The length of a word, or put differently, the number of syllables is also an affecting factor. See the example in (11).

(11) nekɯsɯtakɯrɯ v ekɯsɯtakɯrɯ
neɾɯza v eɾɯza

Again, the only difference in both pairs is [ne] versus [e]. The examiner judged the first pair to be similar but the second different. From this difference in decision, it is suggested that the total length, or the number of syllables of a word can be a decisive factor. It is reasonable to think that the longer a word is, the more the difference is apt to be obliterated in perception. To summarize this section, similarity judgment can be affected at least by the following three factors.

(12) 1 Segmental contents of the nucleus vowel (and possibly the onset consonant)
2 Position of the syllable in a word
3 Length of the word or the number syllables composing the word

Thus the decision is a result of an interaction of these factors.

Finally, let us consider an interesting example in (13).

(13) ɰaiesɯesɯ v aiesɯesɯ
ɰaikeːkeː v aikeːkeː

The difference in phonological contents of the onset and nucleus, the position of the syllable concerned in the word and the length of the word are exactly the same in (13). However, the first pair was decided to be similar but the second different. This is an obvious case of inconsistent decision. This strongly suggests that the similarity decision can depend on the examiners, because, as has been discussed above, perceptual

impression can be different across listeners. However, this does not necessarily deny the possibility of other factors at work. Further inquiry is required in order to bring to light such factors, if any.

6. Conclusion

The present study surveyed how the proposed trademarks are examined in Japan, focusing on trademarks applied from abroad. Under the Japanese trademark law, proposed trademarks are demanded to be transliterated in *Katakana*, which may distort their original phonological structures. This may end up with the application being rejected. It is also discussed that he decision of phonetic similarity is affected by certain phonological and cognitive variables. We have also seen different decisions can be made depending on who examines the case. It is not realistic to demand of examiners, judges and lawyers to be armed with linguistic knowledge. If this is the case, they should consult experts in linguistics. As Dinnsen (1987) claims, trademark law draws crucially on the assessment of phonetic similarity and likelihood of confusion. Phonetics and linguistics thus find applications in law.

Acknowledgements

I wish to extend my sincere gratitude to Biljana Čubrović for her questions, comments, and encouragement. Thanks also go to Patricia Ashby and Hiroko Saito, whose questions and comments helped improve the first version. This work is supported in part by grants-in-aid from the Japan Society for the Promotion of Science (20540353 and 23320083).

References

Bladon, R. A. W. & B. Lindblom. 1981. Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America* 69: 1414–1422.

Dinnsen, D. 1987. Phonetic similarity and trademark law. *Innovations in Linguistic Education* 5.1: 3–9.

Enuenu-kikaku 2003. *Shohyo Shinketsu Shoko Ruihi Shu 2003* (*Database of Sound Similarity of Trademarks in Appeal Decisions: 2003 version*)

Hall, M. 2004. The mystery of similarity. *Managing Intellectual Property*, 137. Japan Special Focus: vi–viii.

Hood, J. D. and J. P. Poole. 1980. Influence of the speaker and other factors affecting speech intelligibility. *Audiology* 19: 434–455.

Ogawara, M. H. 1999. A trademark dispute in the Japanese Court – the Snack Chanel case. Paper presented at the *International Association of Forensic Linguistics*, Birmingham.

Pickett, J. M. 1957. Perception of vowels heard in noises of various spectra. *Journal of the Acoustical Society of America* 29: 613–620.

Sakata, T. 2008. The phonetics of trademark similarity. Unpublished BA Thesis, Osaka University of Foreign Studies.

Shuy, R. 2002. *Linguistic Battles in Trademark Disputes*. New York: Palgrave.

Tabata, T. 1993. Gairaigo-ni kansuru oninseiyaku-ni tsuite. *On-in Kenkyu: Riron-to Jissen*:133–136. Tokyo: Kaitakusha.

PART IV:

PHONOLOGY AND BEYOND

LEVEL ORDERING OF AFFIXES: A PHONOLOGICAL PERSPECTIVE

JELENA VUJIĆ

Outline

Although some theoreticians have tried to ignore the phonological aspect of the mechanisms that are responsible for the affix combinations in English (Fabb 1988), it is clear that when it comes to level ordering and restrictions that govern affix combinations in English, phonology and morphology strongly interact and are at a very intricate interplay. In this chapter we will give an outline of various theoretical approaches to English word-formation from TG (Chomsky and Halle 1968), to a very influential work of Siegel (1974), to The Lexical Phonology and Morphology model (Kiparsky 1982), to Giegerich (1995), and finally, to the most recent theories such as Optimality Theory (Raffelsiefen 2004), all of which consider and acknowledge different aspects of the interdependency that exists between phonology and word-formation. We will show that whichever aspect of such interactions is described and whatever model is used for the purpose, the relation between word-formation and phonology is complex and multi-aspectual. We will show how they can both restrict and aid one another.

1. Introduction

Ordering of affixes is one of the most intriguing topics in English word-formation. Scholars have been trying to fathom the exact mechanisms and features that restrict seemingly perfectly logical and acceptable affix combinations and discard them as ill-formed.

In efforts to shed more light on the combinatorial properties of derivational suffixes one must acknowledge the importance of phonology-morphology interface. That some morphological mechanisms and patterns have effect on phonological rules, just as some phonological patterns affect the morphological choices is nothing new, and it seems only logical

that a large portion of responsibility for regulating suffix combinations in English lies within this constant morpho-phonological interaction.

One of the most striking aspects of the interface between word-formation (WF) and phonology is seen in the stress system and the shift of stress which occurs as a result of WF. English stress is known to be sensitive to the following:

a) **Lexical category of the word being stressed.** Verbs are stressed on the last syllable if it is heavy (1). Otherwise, their stress lies on the penult (2).

(1) obe**l**y, ato**l**ne, arri**l**ve, mole**l**st, colla**l**pse
(2) asto**l**nish, e**l**dit, deve**l**lop

b) **Morphological structure of the word (simple/complex words).** For example, adjectives are typically stressed as verbs (final/penultimate system) if they are unsuffixed, but follow the noun pattern (penult/antepenult) if they are suffixed.

(3) innoc+ence/ innoc+ ent; signific+ance/ signific+ant

c) **Type of affix added.** It is not the same, however, which affix is chosen or added to the word since the stress allocation depends on the choice of affixes. Some affixes are relevant for the stress of a derivative (4a), while the others are stress-neutral (4b).

(4a) adventure > adventurOUS adjective > adjectivAL
(4b) move > moveABLE boy> boyHOOD

d) **Whether the word is related to an independently occurring word.**

2. Types of affixes

While neutral affixes do not interact in the basic syllabification of a word and fall outside the phonological word, which is illustrated in (4b), those presented in (4a) play active part in the basic syllabification of a word and induce changes in the stress pattern of a derivative.

The two affixal (suffixal) types have been referred to differently within different theoretical frameworks. Despite different terminology, they more or less describe the same morpho-phonological phenomena.Whitney (1889), in his description of Sanskrit, refers to them as *primary* and *secondary*. Chomsky and Halle (1968) call them *stress-affecting* and *stress-neutral*. Siegel (1974) defines stress-affecting affixes as *class 1* and those affixes that do not interfere with the stress of a derivative as *class 2*. Within the framework of Lexical Phonology and Morphology they are

referred to as *level 1* and *level 2*, while modern linguistic theories operate with the names *cohering* and *non-cohering affixes.*

In English non-cohering (neutral) are known to be the following suffixes: *-able, -er, -en, -ful, -hood, -ish, -ism,- ist, -ize, -less, -like, -ment, -ness, -ly, -wise, -y* (adjective-forming), and all inflectional suffixes.

Cohering (stress-affecting) properties manifest the following English suffixes*: -age, -al, -an, -ant, -ance, -ary, -ate, -ic, -ion, -ify, -ity, -ory, -ous, -ous, -y* (noun-forming).

3. Phonology-morphology interface in linguistics

a. Chomsky and Halle (1968) and the subsequent theories

Ever since Bloomfield (1933) it has been assumed that there are two kinds of suffixation processes in English, characterizable by a number of differences with respect to their morpho-phonological behaviour. Although many linguists had noticed the phonology-morphology interaction before, it was not until Chomsky and Halle published their *The Sound Pattern of English* (1968) that a comprehensive insight into sound and stress system of English was provided. The study is written from the perspective of transformational grammar, therefore morphology is not allowed much interference. According to TG, morphological rules do not have access to derived phonological information such as the position of the stress. In other words, lexical phonology is segregated from post-lexical one. The interaction of phonology and word-formation in this framework is reduced to be as simple as: word-formation assembles all morphemes and phonology deals with the result. For not recognizing the relation between morphology and phonology in word-formation, such an approach has often been called *a non-interactionist model*. According to this approach, the distinction between two types of suffixes is associated with the following two kinds of boundaries to which they attach:

a) the word or strong boundary #, and

b) the morpheme or weak boundary +.

Affixes marked # attach to words, or more precisely, they are outside the domain of cyclic phonological rules, such as stress assignment. Affixes marked + attach to other morphemes, and they themselves do not block stress assignment.

An often cited example is (5) taken from Plag (1996: 770):

(5) atolm > atolm+ic > atolm+ic+ity

atolm > atolm#less > atolm#less#ness

The Sound Pattern of English gave rise to a number of other theories, and theoreticians who worked on Halle and Chomsky's ideas further elaborated and alternated them. One of them was Dorothy Siegel (1974) who argues that each affix is associated with only one boundary. This approach can be considered as an *interactionist model.* From this argument of hers, Selkirk (1982) later established Affix ordering generalization. According to this principle, weak boundary affixes + (class I) are always attached before, and strong boundary affixes # (class II) are attached after the stress assignment.

b. Lexical Phonology and Morphology

Other researchers and scholars were inspired to work on the ideas proposed within Siegel's interactionist model, which gave rise to one of the most widely adopted views of morphology-phonology interaction and that was Lexical Phonology and Morphology (LPM) model (Kiparsky 1982; Mohanan 1982; Hargus and Kaisse 1993). They unified apparent generalizations about phonology of cohering suffixes and their interaction with morphology (especially their triggering cyclic rule application) and their linear order. They elaborated on Siegel's claims and developed a *level ordering hypothesis.* The hypothesis states that cohering, stress affecting (+ boundary) affixes or *level I* affixes in LPM occur closer to the root, while non-cohering, stress-neutral (#boundary) affixes, or *level II* affixes in LPM, always occur outside the cohering ones. In other words, level I affixes do not attach to the bases already affixed with level II affixes. Thus, level II affixes close a word for further derivation. The reason why some seemingly possible combinations of affixes are blocked lies in the fact that the words that serve as derivational bases contain a non-cohering suffix which blocks further affixation, which is illustrated in (6):

(6) *happi#ness+al *sing#er+ous

BUT

person+al+ity danger+ous#ness (Kaisse 2005: 35)

Level I suffixation followed by level II is acceptable and legitimate just like a suffixation with a string of level II affixes as in seem#less#ly, interest#ed#ness, thought#less#ness.

According to LPM, morphological operations occur one affix at a time. When it comes to the combinability of level I affixes, the intervention of phonology and morphology occurs cyclically in the following manner: when a level I affix is added, the form is passed to level one phonology which involves, among others, processes such as stress assignment rules, trisyllabic laxing, velar softening, sonorant syllabification, etc. The form then returns back to the morphology on that same level, and then back to phonology, and so on. This repeats until all level I affixes for that particular word are added. It seems that in English this cyclic rule application happens only at level I.

LPM model was probably the last phonology-morphology model to receive a wide consensus. However, it appears to have demonstrated its flaws, at least when it comes to English. It is known that affix ordering in some cases proves to be too strong, which is illustrated in (7). If affix ordering rules were to be followed, the combination of *-ment* and *-al* in that exact order would not be possible, and yet, it is not only an acceptable but also a perfectly grammatical combination which as a result creates well-formed derivatives.

(7) #ment+al as in governmental, developmental, argumental (Kaisse 2005: 36)

Even some of the LPM followers acknowledged the too committing and binding nature of level-ordering hypothesis and suggested certain modification. Halle and Mohanan (1985: 64) used the so called LOOP device which "allows a stratum distinction for the purposes of phonology, without imposing a corresponding distinction in morphological distribution".

c. Fabb's criticism of level-ordering hypothesis and the alternatives to Fabb's approach

One of the strongest criticisms of level ordering and arguments against this hypothesis was given by Fabb (1988). He performed a thorough research and analysis of 43 common English affixes. By applying level-ordering hypothesis we could expect some 459 combinations, while English words contain only about 50 pairs of suffixes. The reason for that lies in the fact that some 28 suffixes never combine with the other suffixes. Others show partial or rather restricted productiveness and combinability. From his finds Fabb puts forward the hypothesis that suffixation is constrained by selectional restrictions (e.g. part-of-speech, type of

derivational base since affixes are sensitive to word-internal structure, the origin of derivational base, etc.) of the affixes involved. These selectional restrictions were offered as a serious alternative to level-ordering.

Although Fabb's claims seemed plausible and strong enough at the time, they were seriously challenged by Plag (1996). In his response to Fabb with a series of counterarguments, Plag provided strong support to the claim that the failures of certain suffixes to attach to already suffixed forms are the natural consequence of phonological, morphological, semantic and pragmatic constraints. He also reproached Fabb for generalizations such as "affixes that attach outside one other suffix" or "affixes that do not attach outside any affix", saying that they have no theoretical value. Additionally, Plag supports Giegerich's (1995) arguments for the base-drivenness of morphological processes, where the ability of a particular affix to attach to certain class of stems rests upon the properties of the base rather than the properties of the suffix itself.

In her works Hay (2000, 2002) proposes an account of ordering based on parsability called *complexity based ordering*. Hay's summary of the main insights of Lexical Phonology is that affixes create different boundary strengths and that boundary strength is related to ordering. This boundary strength is seen as gradient rather than stratal (a view similar to that found in Optimality Theory). She provides a psycho-linguistic explanation for a suffix's capacity for combinability. The less phonologically segmentable, the less transparent, the less frequent and the less productive a suffix is, the more resistant it will be to attaching to already affixed words. Furthermore, a particular suffix has a range of separability. In some words it is more separable than in others.

In their co-authored paper Plag and Hay (2004) further elaborate on what constrains possible suffix combinations including structural (selectional restrictions which are mostly grammatical by nature) and psycholinguistic (processing constraints) aspects.

d. Recent developments

Although much challenged and argued, LPM model continues to survive in a somewhat changed form, or as Kaisse (2005: 38) puts it "married to Optimality Theory". The followers use ranked and violable constraints in conjunction with a division among stem-level, word level, and postlexical strata. Constraints can be ranked differently at each stratum, and the output of each stratum is used as the input to the next.

There is no doubt that the level ordering showed many weaknesses. Explaining the nature of strata as an epiphenomenon of underlying phonological properties of suffixes of English weakens the idea of strata. Raffelsiefen who works within Optimality Theory shows that not even two of the many suffixes of English trigger exactly the same type of morpho-phonological alternations, so that we would need as many sub-strata as we have suffixes that trigger morpho-phonological alternations. Therefore, it is better to think of a continuum of strata than simple dipartite system. (Raffelsiefen 1999; Hay & Plag 2004: 568). She further challenges the division of suffixes in two groups only (Raffelsiefen 2004). Her equivalent to a coherent grouping of stem-level versus word-level affixes is a group of affixes which cause the forms to which they attach to be evaluated by one ranked set of constraints versus a second set which call up another ranking of those constraints.

Such behaviour is manifested by rival suffixes *-ize,* which is neutral, and *-ify*, which possesses stress-changing properties (8) and (8b).

(8a) *Bushize/ Bushify/ Clintonize

(8b) carbonify (to convert to carbon, to burn or caramelize) vs.
carbonize (to unite with carbon, to enrich or coat with carbon)

(8c) ozonize = ozonify

(8d) ionize/ ionify*

Raffelsiefen shows that both affixes require a source to which to match their stressed syllables. Their meaning is the same, denoting denominal verbs and both suffixes are used to form verbs from nouns/adjectives. For that reason, they are considered rival suffixes. In (8b) suffixes *-ize* and *-ify* are used to attribute different meaning to the verbs derived from the same base. In (8c) both suffixes create denominal verbs with the same meaning, while in (8d) for some reason, there is a kind of restriction at play which does not allow the suffix *-ize* to be applied, though this combination is seemingly possible. Raffelsiefen argues that the choice of the two suffixes in question largely depends on a speaker's lexicon (the reason why in some cases both can be attached to the same base with the same meaning). Suffixes *-ize* and *-ify* attach to suitable stems in the speaker's lexicon to avoid ill-formed phonological results. The given affixes can also be considered cohering (level 1) in the way that they are phonologically interactive with their bases.

The pre-existence of phonologically suitable word which is necessary to permit a new coinage, is a notion that Raffelsiefen borrowed from Steriade (2000). Steriade claims that derivational morphology is powerful enough to coerce a derived word to agree in phonetic details with other members of its paradigm. She further explains the interaction between morphology and phonology explaining the similarities that exist in phonetic features of the derivative and the base. She argues that such similarities could be the result of the phonology acting in boundary delimiting function, or in other words helping mark beginnings and end of morphemes.

4. Conclusion

This paper presents a tour through the various aspects of phonology-morphology interactions and their implication to suffixation processes. Although sufficiently explanatory for other languages, the stratum model obviously has both its advantages and disadvantages as it fails to account for many phenomena when it comes to mechanisms that regulate suffix combinations in English. The gliding continuum model offered within the scope of Optimality Theory, and some recent studies (mostly done by morphologists, such as the advocates of Distributed Morphology theory) show that phonology and morphology are locked in a tight inseparable relation when it comes to derivation in English.

References

Chomsky, N. & M. Halle. 1968. *Sound Pattern of English.* New York: Harper & Low.

Fabb, N. 1988. English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory* 6: 527–539.

Giegerich, H. 1995. *Lexical strata in English: morphological causes, phonological effects.* Cambridge: CUP.

Halle, M. & K. P. Mohanan. 1985. Segmental Phonology of Modern English. *Linguistic Inquiry* 16: 57–116.

Hargus, S. & E. Kaisse (eds.). 1993. *Studies in Lexical Phonology.* San Diego: Academic Press.

Hay, J. 2000. *Causes and Consequences of Word Structure.* Unpublished Doctoral Dissertation, Northwestern University.

—. 2002. From Speech Perception to Morphology: Affix-ordering Revisited. *Language* 78 (3): 527–555.

Hay, J. & I. Plag. 2004. What constraints possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory* 22: 565–597.

Kaisse, E. 2005. Word-formation and phonology. In *A Handbook of Word-formation,* edited by P. Stekauer & R. Lieber, 25–47. Dodrecht: Springer.

Kiparsky, P. 1982. From cyclic phonology to lexical phonology. In *The structure of phonological representations, Pt.I,* edited by H. Van Den Hulst & N. Smith, 131–175. Dodrecht: Foris.

Mohanan, K. P. 1982. *Lexical phonology.* Unpublished Doctoral Dissertation, Massachusetts Institute of Technology.

Plag, I. 1996. Selectional restrictions revisited: a reply to Fabb (1988). *Linguistics* 34: 769–798.

—. 1999. *Morphological Productivity, Structural Constraints in English Derivation.* Berlin/New York: Mouton de Gruyter.

Raffelsiefen, R. 1999. Phonological constraints on English word-formation. In *Yearbook of Morphology 1998*, edited by G. Booj & J. van Marle (Hrsg.), 225–287. Dordrecht: Kluwer.

—. 2004. Absolute ill-formedness and other morphological effects. *Phonology* 21: 91–142.

Selkirk, E. 1982. *The Syntax of Words.* Cambridge, MA: MIT Press.

Siegel, D. 1974. *Topics in English morphology.* New York: Garland.

Steriade, D. 2000. Paradigm uniformity and the phonetics-phonology boundary. In *Papers in laboratory phonology V. Acquisition and the Lexicon,* edited by M. Broe & J. Pierrehumbert, 313–334. Cambridge: CUP.

Whitney, W. D. 1889. *Sanskrit grammar.* Cambridge, MA: Harvard UP.

THE FUNCTIONAL CLASSIFICATION OF ENGLISH VOWELS: PHONOLOGICAL AND ORTHOGRAPHIC EVIDENCE

CSABA CSIDES

Outline

The aim of this article is to demonstrate that the division of English vowels into tense and lax is based on phonological evidence as well as orthographic justification. Some authors use these categories as phonetic labels and generally claim that tense vowels are produced with more tension of the articulatory muscles, length and diphthongization than the lax vowels. In the first part of the article, I would like to argue that phonological alternations seem to support the view that the categories of tense and lax are indeed functional/phonological in nature. Phonological processes discussed in connection with these arguments are Vowel Shift, Trisyllabic Laxness, Laxing by ending, CiV laxing, Pre-cluster laxing and Laxing by free U. The next section of the article concentrates on the regular sound values of English vowel letters and discusses the difference between free and covered graphic positions. It turns out from orthographic evidence that tense and lax vowels respectively tend to occur in different types of graphic (orthographic) positions in the default case. The effect of the free position rule, however, may be eliminated by stronger (overriding) regularities of the language that turn out to be phonological in nature.

1. Introduction

Traditionally vowels are classified phonetically along several dimensions in the literature. For example, length (quantity) is usually recognized as a feature useful in the classification of vowels, and it is claimed that the pronunciation of a long vowel takes up roughly twice as much time as the pronunciation of a short vowel. Furthermore, several articulatory gestures are involved in the phonetic classification of vowels. The rounded-unrounded, monophthong-diphthong division and the tongue

body (quality) features like [high], [back], [low] are all familiar to those somewhat versed in phonetics.

The stressed-unstressed, strong-weak, full-reduced distinctions are also regarded as phonetic categories by some linguists although it is more difficult to characterise them in articulatory terms. Stress is usually regarded as a relational term and according to some phoneticians it more easily lends itself to an acoustic characterisation than to an articulatory one and it is intimately connected to vowel reduction.

When it comes to the tense-lax dichotomy, we notice that these features are a source of uneasiness amongst phoneticians and phonologists. The trouble is caused by the fact that these features are sometimes used as phonetic categories, sometimes as functional labels. When used as phonetic labels, it is usually assumed that these features refer to the relative tension of the articulatory muscles during the production of a sound. Kreidler (1989), among others, uses these features as articulatory labels describing the relative intensity in the articulatory production of the sound.

Some authors simply put an equation mark between [long] and [tense] on the one hand, and [short] and [lax] on the other. For example, Katamba (1993: 48) claims that "the English 'long' vowels and diphthongs i, ɑ, ɔ, u, eɪ, aɪ, ɔɪ, əʊ, juˈ are tense, while the 'short' vowels [ɪ, e, æ, ʊ, ʌ, ɒ] are lax." Durand (1990: 53) claims that [+/- tense] "has proved to be one of the most controversial features in the history of phonology. Whereas specialists have often been able to accept its validity for consonants, in the case of vowels it has been felt that the evidence for muscular tension/laxness was lacking and that [+/- tense] merely conflated other distinctions, which were independently needed such as [+/-long] or [+/- centralized]". A similar claim is made by Giegerich (1992: 98), who finds that "tense vowels are produced with a deliberate, accurate, maximally distinct gesture that involves considerable muscular effort: nontense sounds are produced rapidly and somewhat indistinctly". He (1992: 101) also couches his observation in the form of a 'redundancy rule', given in (1) below.

Table 1. Redundancy rules

[+tense] → [+long]
[-tense] → [-long]

The term 'redundancy rule' implicitly expresses the fact that either the feature [tense] or the feature [long] is superfluous in a framework that

treats tense as a purely articulatory label. The term *redundancy* itself suggests that either the feature [tense] or the feature [long] must be dispensed with as a technical device in phonological descriptions if one attempts to adhere to a fairly constrained model.

In sum, when used as articulatory labels the terms *tense* and *lax* are frequently equated with the long/short dichotomy, and it is claimed that the long vowels are tense because they are pronounced with more intensive tension of the articulatory muscles, they are longer than the lax vowels and they are more susceptible to diphthongization.

The problem with this argumentation is that it directly leads to the violation of a famous scientific principle commonly known as Occam's razor. Occam, the English philosopher, came to the conclusion that in any scientific analysis we must not multiply theoretical entities beyond necessity. And indeed, if there is no difference between [tense] and [long] there is no point in keeping both terms in our technical vocabulary.

2. The phonetic and functional classification of British RP vowels

While the English consonant system is relatively stable across dialects, there are some differences among vowel systems. It is vital, therefore, that we specify the dialect when we classify the vowels of 'English' in order to know whereof we speak. In what follows we will concentrate on the vowel system of standard southern British English, technically known as Received Pronunciation (RP). Table (2) shows the phonetic classification of BrE RP vowels along the short/long division. As it appears from the table, there are 5 long monophthongs and 8 diphthongs in RP, which means that there are altogether 13 long vowels in this dialect. Besides the long vowels there are 7 short monphthongs out of which 6 can function as a full vowel. The only monophthong that appears in parentheses in table (2) is the schwa '/ə/', the prime unstressed vowel of English. Besides the 'schwa', /ɪ / and /ʊ/ can also appear in unstressed syllables.

Table 2. Quantity in RP

Long vowels	Short vowels
/iː/, /uː/, /ɔː/, /ɜː/, /ɑː/, /aɪ/, /aʊ/, /ɔɪ/, /eɪ/, /əʊ/, /ɪə/, /ʊə/, /eə/	/ɪ/, /ʊ/, /e/, /ʌ/, /æ/, /ɒ/, (/ə/)

The classification in (2) can be justified on phonetic and also on phonological (distributional) grounds. Long vowels are excluded from certain phonological position and vice versa: there are some phonological contexts in which short vowels may not occur. One of the well-known facts about English words is that they may not end in a short stressed vowel, for example.

When compared to the traditional phonetic classification, the tense/lax division gives a different result. As noted above in the introduction, we must treat the terms 'tense' and 'lax' as functional/phonological labels in order to be able to retain them in our technical parlance.

Table 3. The functional classification of vowels in BrE RP

Tense vowels	Lax vowels
/aɪ/, /aʊ/, /ɔɪ/, /eɪ/, /əʊ/, /ɪə/, /ʊə/, /eə/ /iː/, /uː/, /ɔː/	/ɪ/, /e/, /ʌ/, /æ/, /ʊ/, /ɒ/, /ɜː/, /ɑː/, /ɔː/

It appears from table (3) that all the diphthongs are analysed as tense vowels while all the short monophthongs behave as lax vowels. The long monophthongs, however, are distributed between the two categories: the high long moniphthongs are [tense] whereas the non-high long monophthongs are analysed as [lax]. The only problematic item is /ɔː/, which shows a doubled-faced behaviour. It is for historical reasons that /ɔː/ is sometimes analysed as [tense], sometimes as [lax], and this problem will not be discussed in the present paper.[1]

When comparing table (2) to table (3) above the attentive reader immediately noitices the fact that long vs. tense and short vs. lax simply do not coincide. There are, of course, overlapping elements but the respective categories just do not coincide. The proposal of the present paper is that it is possible to use the tense/lax categories as additional labels for classifying British RP vowels without multiplying categories beyond necessity because some long vowels seem to pattern together with short vowels while others line up with long vowels. High long vowels pattern together with diphthongs, while non-high long vowels pattern

[1] The interested reader is invited to consult Wells (1982) concerning the historical developments that resulted in present day RP /ɔː/.

together with short monophthongs, /ɔː/ is sometimes analysed as tense, sometimes as lax. In section (3) below we have collected some phonological evidence in favour of the classification in (3) above.

3. Regularities supporting the functional classification

This section is devoted to regularities, alternations and distributional facts that seem to support the view that the terms [tense] and [lax] may be used as additional labels for classifying English vowels without introducing unnecessary categories into our technical jargon.

3.1. Trisyllabic Laxness

Trisyllabic laxness is an amply documented phenomenon in textbooks dealing with the phonology of English. According to the received wisdom, only lax vowels tend to occur in the antepenultimate stressed syllable of English words unless there is an internal analytic (word-level) boundary inside the word. The phenomenon thus is a structure dependent regularity whose operation is dependent on the internal morphological composition of the word. For this reason this regularity is regarded as a lexical rule in some theoretical frameworks, but it is also frequently referred to as a morpho-phonological regularity. If the word has a synthetic (monolithic) structure the regularity freely operates. If, however, an internal analytic (#) boundary cuts the word into two independent phonological domains the rule breaks down. Consider now the synthetic forms in table (4) all obeying Trisyllabic Laxness.

Table 4. Trisyllabic Laxness

/æ/	/e/	/ɪ/	/ɒ/
cápital	rélevant	divínity	persónify
vánity	sevérity	críminal	hóliday
sácrament	compétitor	míracle	ferócity
bálustrade	régiment	ínstitute	hólograph
animal	pénalty	prímitive	ánimosity
grávity	élegance	pýramid	órigin
mánager	élement	vírulent	cónifer

/ʌ/	/ɑː/	/ɜː/	/ɔː/
lúxury	pártisan	círcular	córdial
ultimate	párliament	túrbulence	dórmitory
adúltery	ártery	intérpolate	inórdinate
núnnery	márvellous	fértilize	páucity
cómpany	márginal	términate	fórmulate
súmmary	lárceny	pérsecute	córpulent

Table (4) above contains examples of both static and dynamic Trisyllabic Laxness. In other words, there are items in which TSL is manifest in a static manner. For example, *capital*, *relevant*, *luxury* and *animal* have no shorter stem alternants with a tense vowel, while *vanity, sacrament, divinity* and *criminal* have, cf. *vain, sacred, divine* and *crime*. In the latter four examples we can talk about stem-vowel alternation or vowel-shift where Trisyllabic Laxness manifests itself in dynamic manner. This alternation, however, is heavily dependent on the internal structure of the given word as it is evidenced by the examples in (5) below.

Table 5. Analytic and non-analytic domains

No Trisyllabic Laxness		**Trisyllabic Laxness**	
tidy	tidiness	penal	penalty
lazy	laziness	compete	competitive
fever	feverish	creed	credible
demon	demonish	severe	severity
meagre	meagrely	hero	heroin
final	finally	divine	divinity
total	totally	derive	derivative
legal	legally	crime	criminal
spiteful	spitefully	tyrant	tyranny
maiden	maidenhood	grave	gravity
parent	parenthood	grade	graduate
advise	advisable	insane	insanity
read	readable	nation	national
reason	reasonable	fable	fabulous
bribe	bribery	prepare	preparatory
note	notary	legal	legislate
bounty	bountiful	sole	solitude

leader	leadership	mode	modify
dictator	dictatorship	holy	holiday
cater	catering	compose	compositor

The pairs of lexical items appearing in the first two columns of (5) do not show tense-lax stem-vowel alternation, i.e., TSL seems to break down in these cases. Notice, however that in each of these examples there is an analytic boundary between the stem and the suffix. These suffixes are attached to the stem with an analytic boundary and when we remove them we still end up having a free stem, an independently pronounceable meaningful word. This is why the analytic boundary is also referred to as a 'word-level boundary'. The suffixes themselves are also fairly independent in the sense that they are meaningful and relatively productive. The situation in the last two columns is radically different in that there is tense-lax stem vowel alternation in the antepenultimate stressed syllable of these words. The items in the last column, however, are structured in a different way, if structured at all. The 'suffixes' appearing in this column are not real suffixes in the sense that they cannot be used independently for word-formation, they are thus not productive and their meaning is rather vague in the vast majority of the cases. They are thus endings appearing at the end of a restricted set of lexical items rather than real suffixes. These strange, fossilized entities have been referred to in the literature as 'synthetic suffixes', and the items have been termed root-derived forms, cf. Harris (1994), for example. Note also that the stems are not free stems in these cases, they cannot be used independently in an utterance, they are regarded as bound stems or roots, hence the term root-derived forms is frequently used by some authors..

Since Trisyllabic Laxness is blocked by a word-level boundary the regularity is regarded as a structure dependent 'rule' of the language. Although there are a number of exceptions to TSL in English both systematic and irregular[2], it still remains true that tense vowels in general

[2] The most famous systematic exception to TSL is the case of the vowel /(j)u:/, which seems to occur freely in stressed antepenults. Think of *purify, opportunity, communicate, musical, endurance, souvenir, funeral, jubilee, unity, mutiny, curious, putative, lucrative* etc., for instance. Moreover, words ending in *–ery, –ary, -ory* also seem to defy the regularity, cf., *ivory bribery, plenary, notary, vagary, library, binary, scenery*, etc. The stressed vowel letters O, U, A or E

seem to escape the antepenultimate stressed syllable of English words in lexical items having monolithic structure.

3.2. Pre-R Breaking and Pre-R Broadening

The rule of Pre-R Breaking and Pre-R Broadening are also amply documented in the literature, cf., for example, (Wells 1982, Nádasdy 2006, Csides, in prep). Both of these regularities are the end product of historical developments in the English language and they have caused the restructuring of the English vowel system. Since the discussion of the diachronic steps of these phenomena would take us far beyond the scope of the present study, we will only concentrate on the synchronic distribution of English RP vowels in the pre-r context. Readers interested in the historical background of these processes should consult Wells (1982).

3.2.1. Pre-R Breaking of British RP

It is a commonplace synchronic phonological statement about British English RP that the Plain Tense vowels may not occur before historical r and they are replaced by the set of Broken Tense vowels. Table (6) below list the Plain Tense-Broken Tense pairs of British English RP vowels.

It appears from table (6) that the high long vowels and the closing diphthongs have been replaced by centring diphthongs/triphthongs[3] in the history of this dialect. The only monophthong among the Broken Tense vowels is the vowel /ɔ:/, which also used to be pronounced /ɔə/ before undergoing monophthongization. This distribution is also true in cases where the /r/ remains ultimately unpronounced as a result of r-dropping.

---

followed by a (*C*)onsonant-letter, letter *i*, (*V*)owel-letter sequence also resist TSL as in *appreciate, medium, senior, mediate, radiate, podium*, etc. Some compound like items of foreign origin also escape the regularity, think of *patriarch, micrograph, prototype*. Besides these systematic exceptions we may also come across a handful of unsystematic exceptions that must be memorized one-by-one, *stabilize, patriot, nightingale, favourite, irony* and a handful of others.

[3] Note that not all the phonologists recognize the existence of triphthongs. Some claim that triphthongs are split into two syllables. This issue is not investigated here.

Consider the examples in (7) below where the historical /r/ remains unpronounced in all the items of the broken tense column.

Table 6. Pre-R Breaking

	Plain tense	*Broken tense*
Monophthongs	/iː/	/ɪə/
	/(j)uː/	/(j)ʊə/
Diphthongs	/aɪ/	/aɪə/
	/eɪ/	/eə/
	/əʊ/	/ɔː/
	/aʊ/	/aʊə/
	/ɔɪ/	/ɔɪə/

Table 7. Pre-R Breaking with examples

	Plain tense	*Broken tense*
Monophthongs	/iː/ *been*	/ɪə/ *beer*
	/(j)uː/ *moon*	/(j)ʊə/ *moor*
Diphthongs	/aɪ/ *time*	/aɪə/ *tyre*
	/eɪ/ *stain*	/eə/ *stair*
	/əʊ/ *stone*	/ɔː/ *store*
	/aʊ/ *house*	/aʊə/ *hour*
	/ɔɪ/ *join*	/ɔɪə/ *Moir*

What is interesting about Pre-R Breaking is that it has only affected the tense vowels of English and also that it has led to a skewed distribution of tense vowels. In a pre-r environment only the broken variety may occur even if the /r/ remains ultimately unpronounced. It is worth bearing in mind that Pre-R breaking similarly to Trisyllabic Laxness is a structure dependent regularity of the language in as much as it breaks down across an analytic boundary, cf. *re#*read is pronounced /ˌriːˈriːd/ but never */ˌrɪəˈriːd/ or *low#risk* is always pronounced /ˌləʊˈrɪsk/ and never */ˌlɔːˈrɪsk/.

3.2.2. Pre-R Broadening (Compensatory Lengthening)

Another major restructuring of the English vowel system affected only the set of lax vowels. The short 'plain' lax vowels have undergone lengthening as a result of r-dropping. This phonological restructuring is

known technically as ‘compensatory’ lengthening and it is regulated by the principle of phonological quantity preservation. This type of lengthening is called ‘compensatory’ because the lengthening of the vowel compensates for the loss of the historical r. This is a widespread process cross-linguistically: in a number of languages a timing slot has been vacated by the loss of a consonant from before another consonant and this empty timing slot has been targeted by the preceding vowel, which was originally short. By spreading into the vacant timing slot the vowel becomes long and it thus compensates for the deletion of the consonant preserving the original overall quantity of the lexical item.

It is also interesting to note that the non-low plain lax vowels have undergone merger in this phonological context resulting in a mid-central long vowel /ɜː/, informally referred to as the ‘long schwa’. The Plain Lax – Broad Lax pairs of vowels are given in table (8) below.

Table 8. Pre-R Broadening

	Plain lax	*Broad lax*
Low vowels	/æ/	/ɑː/
	/ɒ/	/ɔː/
Non-low vowels	/ɪ/	/ɜː/
	/e/	
	/ʌ/	
	/ʊ/	

Since – as mentioned above – broadening is a case of compensatory lengthening no /r/ is pronounced after the broad-lax vowel in these lexical items even though there is one in the spelling. It is also straightforward that the /r/ that has disappeared from these lexical items was always followed by a consonant or the end of the word (pause). Consider the words in (9) below having plain lax and broad lax vowels respectively.

Table 9. Pre-R Broadening with examples

Plain lax	*Broad lax*
/æ/	/ɑː/
band, bat, fat, man, bank	bard, bar, far, mar, bark
/ɒ/	/ɔː/
not, fond, pond, conk	nor, ford, port, cork

/ɪ/	/ɜː/
gist, flint, splint, film	girl, flirt, skirt, firm
/e/	/ɜː/
hen, belt, help	her, Bert, herb
/ʌ/	/ɜː/
hunt, cup, fun, pun	hurt, cur, fur, purr
/ʊ/	/ɜː/
push, put, bush	purr, burn

In order to illustrate the point further, consider table (10) below, containing further examples for broadening.

Table 10. Lexical items with broad lax vowels

/ɑː/	park, car, start, pardon, marvellous, mark, star, farther, partner, bar
/ɔː/	port, important, sport, north, fortress, orthodox, boredom, lord, ford
/ɜː/	term, hermit, permanent, her, termite, ternary, terminate, kernel,
	girl, stir, gird, bird, birch, skirt, shirt, skirmish, firm, sir, sirloin, mirth
	murky, burn, hurt, fur, cur, lurch, turn, furnace, turtle, burden, lurk

The fact that the 'Broad Lax' vowels are the result of compensatory lengthening does not mean that all the three Broad Lax vowels come exclusively from this source. In fact, it is only the vowel /ɜː/ whose only source is broadening.[4] The other two Broad-Lax vowels, /ɑː/ and /ɔː/ also come from other sources, cf. words such as *spa, bra, palm, law, bald, pawn* that have never contained an /r/ in their history and probably never will.

It is also worthy of note here that no broadening has taken place when the historical /r/ was immediately followed by a vowel. It is a natural consequence of the fact that /r/ was never deleted from before a vowel and

[4] With the exception of *colonel* whose pronunciation is [ˈkɜːnəl].

hence no timing unit was vacated in these lexical items. Since no timing slot was vacated the preceding plain lax vowel has never had the opportunity to lengthen into the following empty timing unit as there was none available. As a result the Plain Lax vowel remained Plain Lax in these cases. This phenomenon is nicknamed the 'Carrot Rule' by Nádasdy (2006) and it is illustrated in (11) below.

Table 11. Absence of Broadening = Carrot rule

/æ/	*marry, baron, caret, clarify, parrot, carrot, barrel*
/e/	*heron, terror, herring, merit, derelict, ferry, error*
/i/	*myriad, mirror, tyranny, spirit, lyrics, miracle*
/ɒ/	*moral, torrent, sorry, Morrison, foreign, porridge*
/ʌ/	*currant, hurry, worry, Surrey, currency, furrow*

All in all then, the historical developments affecting pre-r vowels support the tense-lax dichotomy since the two sets of vowels have been affected differently by a following historical /r/.

3.3. Vowel-shift: Tense/Lax stem-vowel alternations

English stem-vowel alternations technically known as vowel-shift also support the tense-lax division of English vowels. Interestingly, the problem discussed in this section is the modern reflex of a 15th century restructuring of the English vowel system commonly known as the Great Vowel Shift (GVS). Examples of the four major types of vowel shift are given in (12) below.

Table 12. Vowel-Shift

	Basic type		Pre-r type		
	Tense	Lax	Tense	Lax	Letter
Type I	/eɪ/	/æ/	/eə/	/æ/	A
	sane	sanity	compare	comparative	
	state	stature	barbarian	barbaric	
	valency	valid	declare	declarative	
	grade	gradual			
	fable	fabulous			

Type II	/iː/	/e/	/ɪə/	/e/	E
	creed	credulous	imperial	imperative	
	intervene	intervention	hero	heroin	
	serene	serenity	sphere	spherical	
	compete	competitive	sincere	sincerity	
	creed	credible			
Type III	/aɪ/	/ɪ/	/aɪə/	/ɪ/	I
	derive	derivative	tyrant	tyranny	
	rite	ritual	satire	satirical	
	divide	division	conspire	conspiracy	
	divine	divinity	respire	respiratory	
	final	finish			
Type IV	/əʊ/	/ɒ/	/ɔː/	/ɒ/	O
	code	codify	glory	glorify	
	sclerosis	sclerotic	flora	florist	
	tone	tonic	phantasmagoria	phantasmagorical	
	compose	compositor	folklore	folkloric	
	go	gone			

After having seen the four chief types of vowels shift we must concentrate on the conditioning contexts that trigger the alternation. The first phonological context is the now already familiar Trisyllabic Laxness.

3.3.1. Trisyllabic Laxness

Vowel-shift, i.e. regular tense-lax alternation of stem vowels may be due to Trisyllabic Laxness whereby tense vowels become lax in the antepenultimate stressed syllable of English words unless there is an internal analytic boundary in them. Consider the items in (13) below.

Table 13. Vowel-Shift due to Trisyllabic Laxness

sane	sanity
serene	serenity
derive	derivative
divine	divinity
crime	criminal

mode	modify
compare	comparison
nation	national
compete	competitor
secret	secretary

3.3.2. Laxing endings

There are certain endings that seem to trigger regular vowel-shift in the immediately preceding syllable. This is illustrated by the examples in (14) below. Note that *–ish* seems to function as a laxing ending only in the case of verbs but not in the case of adjectives, for example, cf. *childish, brutish, Danish, stylish* etc., all having a tense vowel before *–ish*.

Table 14. Laxness due to a laxing ending

-ic		-id		-ish (verbal)	
tone	tonic	valency	valid	final	finish
cycle	cyclic	rabies	rabid	rave	ravish
lyre	lyrics			punitive	punish
metre	metric				

3.3.3. Laxness due to a free letter U

Regular tense-lax alternation also takes place in the syllable immediately preceding the vowel-letter U which is situated in a free graphic position. This is a strange characterisation of the environment where vowel-shift takes place since it partly refers to a graphic position. As it will be discussed in section **4.** below, vowel-letters are said to occupy a free graphic position in the cases shown in (15).

Table 15. Free graphic positions

V̲CV	**V̲V**	**V̲#**

The underlined stressed vowel letter is in a free graphic position if it is followed by one consonant-letter and then by another vowel-letter, if it is immediately followed by another vowel-letter or if it occurs directly at the end of the word. In table (15) above capital (V) stands for vowel-letter,

capital (C) for consonant-letter and '#' symbolizes the end of the word. If we examine the items in (16) below, it is immediately obvious that the underlined stressed vowel-letters occurring in the right-hand column are pronounced lax and all of them are followed by the vowel-letter U in the next syllable in a free graphic position. Consider now the examples in (16) below.

Table 16. Laxing due to a free U in the next syllable

case	ca**ua**l
rite	ri**ua**l
grade	gra**ua**l
grade	gra**ua**te
vacant	va**cuo**us
state	sta**ture**
creed	cre**dulo**us
Jesus	Je**sui**t

3.3.4. Laxing due to a consonant cluster (closed syllable shortening)

Another environment in which tense vowels seem to alternate with the corresponding lax vowels is traditionally referred to as closed syllable shortening. Laxing is said to take place before a consonant cluster because in the left-hand column of (17) below the stressed vowel is followed by one consonant and it is tense, while in the right-hand column the stressed vowel is followed by two consonants and it is lax.

Table 17. Laxing due to a consonant cluster

redeem	redemption
intervene	intervention
deceive	deception
keep	kept
weep	wept

3.3.5. iCiV Laxing

The last environment in which vowel-shift takes place can be referred to as iCiV laxing. There are two conditions for iCiV laxing to occur. On the one hand the stressed vowel-letter must be I while on the other hand this vowel must be followed by a (C)onsonant-letter, letter (i) (V)owel-letter sequence, hence the term CiV. Consider the examples in (18) below where the relevant iCiV sequence is emboldened in the right-hand column. This context is held responsible for the regular tense-lax alternation.

Table 18. Vowel-shift due to iCiV laxing

revise	rev**isio**n
divide	div**isio**n
vice	v**icio**us
decide	dec**isio**n
preside	pres**idiu**m

In section 3 we have collected regularities that seem to support the view that the tense-lax division is a viable one when we attempt to characterize the English vowel system. Furthermore, it has probably by now become evident that the recognition of this dichotomy does not lead to an unnecessary multiplication of theoretical tools since the tense-lax division is not the same as he long/short dichotomy.

4. Graphic positions for single vowel-letters

This section is devoted to some orthographic evidence in favour of the tense-lax division. First of all, we will examine the regular sound values for each single vowel-letter and then we will concentrate on the difference between free and covered graphic positions. We will also illustrate the connection between the graphic position and the sound value of the given vowel-letter. Finally, towards the end of the section, we will investigate the regularities that may override the effect of what we will call the *free position rule*.

4.1. Sound values of single vowel-letters

The table in (19) below shows the four regular sound values of each single vowel-letter. It must be noted here that the sound values for each single vowel-letter in the plain tense column correspond to the alphabetical names of the respective vowel-letters. Furthermore, it is also evident that there are only three regular sound values for the vowel-letter O since the Broken-Tense and the Broad-Lax values happen to coincide in this case. Otherwise there are four regular sound values for each vowel-letter, two of which – the Broken-Tense and the Broad-Lax values – are the result of Pre-R Breaking and Pre-R Broadening briefly discussed in section 3 above.

Table 19. Sound values of single vowel letters

Values	Letters	A	E	I=Y	O	U
T E N S E	PLAIN TENSE	/eɪ/ t**a**pe str**a**nge s**a**tan	/iː/ m**e**ter del**e**te l**e**gal	/aɪ/ l**i**ke t**i**ming fl**y**	/əʊ/ st**o**ne b**o**ne al**o**ne	/(j)uː/ f**u**me b**u**gle j**u**niper
	BROKEN TENSE	/eəː/ f**a**re sc**a**rce sc**a**ry	/ɪə/ m**e**re m**a**terial sev**e**re	/aɪə/ l**y**re adm**i**re s**i**ren	/ɔː/ b**o**re st**o**ry **o**ral	/(j)ʊə/ sp**u**rious f**u**ry mat**u**re
L A X	PLAIN LAX	/æ/ h**a**t c**a**rrot c**a**nnon	/e/ b**e**d b**e**ggar l**e**g	/ɪ/ b**i**t m**y**riad b**i**tter	/ɒ/ p**o**t w**o**bble r**o**b	/ʌ/ c**u**t b**u**nny l**u**ck
	BROAD LAX	/ɑː/ b**a**r p**a**rtner d**a**rling	/ɜː/ k**e**rnel h**e**rb m**e**rmaid	/ɜː/ g**i**rl s**i**r b**i**rd	/ɔː/ perf**o**rm **o**rthodox **o**rchard	/ɜː/ m**u**rky l**u**rk t**u**rn

4.2. The graphic position rules

There are two different types of graphic position in which stressed vowel-letters may occur: the free graphic position and the covered graphic position. Note that since we concentrate on the orthographic arrangement following the stressed vowel-letter, we take all letters – even silent ones –

into consideration. This is thus not phonology but letter-to-sound correspondence, and this section serves to illustrate how the English spelling system as a whole indicates the tense or the lax pronunciation of a given vowel-letter in the regular case.

As it appears from table (20) below, there are three different cases of free graphic position and two sub-cases of covered graphic position. In column (a), (b) and (c) the stressed vowel-letters occur in a free-graphic position: in column (a) the stressed vowel-letter is followed by one consonant letter and then by another vowel-letter, V**CV,** in column (b) the stressed vowel-letter is immediately followed by another vowel-letter, V**V**, while in column (c) the stressed vowel letter is at the end of the word. In the examples of column (d) and (e) the stressed vowel-letter is in a covered graphic position. In column (d) the stressed vowel-letter is followed by two consonant-letters, V**CC**, whereas in column (e) the stressed vowel-letter is followed by one consonant but this consonant is immediately at the end of the word.[5]

The attentive reader has noticed by now that the basic connection between the graphic position and the sound value of the given vowel-letter is that vowel-letters are pronounced tense in free graphic positions while they are pronounced lax in covered graphic positions. One can verify this by comparing the sound value of the capitalized stressed vowel-letters in (20) to the regular sound values of each vowel-letter in (19) above.

Table 20. Graphic position rules

Free graphic position			*Covered graphic position*	
(a)	(b)	(c)	(d)	(e)
mA**ke**	chA**o**s	denY**#**	mA**tt**er	hA**t#**
crA**te**r	trU**a**nt	mE**#**	r**Ock**er	h**It#**
dU**ly**	lE**o**	gO**#**	chA**pt**er	cU**p#**
h**Ide**	l**Io**n	flU**#**	stU**mbl**e	flO**p#**
sO**da**	hO**a**x		s**Entence**	scA**n#**
			bU**x**om	
V**CV**	V**V**	V**#**	V**CC**	V**C#**

[5] Note that the consonant-letter *x* counts as two letters for the purposes of calculating the graphic position in column (d). The reason for this is connected to the fact that it represents two consonants [ks].

There are some (apparent) counterexamples to the distribution above. Table (21) contains examples where the capitalized stressed vowel-letter is pronounced tense even though it is followed by two consonant-letters.

Table 21. Apparent counterexamples to the graphic position rules

(a)	(b)
tY**ph**oid	stA**bl**e
hY**ph**en	crA**dl**e
cI**ph**er	scrU**pl**e
scA**th**e	mE**tr**e
pA**th**os	A**pr**on
bA**th**e	cY**cl**e

There are two possible strategies at this point: either we recognize the fact that these are real counterexamples to the generalization concerning letter-to-sound correspondences or we assume that the consonant-letter combinations appearing after the stressed vowel will count as a single consonant-letter for the purposes of defining graphic positions. We will choose the latter strategy, and claim that *ph* and *th* letter-combinations, column (a)[6] and stop plus liquid combinations, column (b) will be calculated as single consonant letters when specifying the graphic position of the stressed vowel-letters of (21). By accepting this proviso we can squeeze the apparent irregularities in (21) into the regular cases of (20) above: they will all belong to column (a) where the stressed vowel-letter is followed by a CV combination.

Although the generalisation about the pronunciation of stressed-vowel letters in a covered graphic position is a fairly powerful regularity in that there are just a handful of – some 40 or so – counterexamples to that, the free position rule may be overridden by stronger regularities already encountered in connection with the discussion of Vowel-Shift in section 3 above.

[6] Recall that the consonant letter *x* counts as two consonant-letters. It must be noted here that the *ph* and *th* digraphs may be regarded as single consonant letters because they represent a single consonant; *ph* usually stands for [f] whereas *th* represents a [θ] or a [ð].

Table 22. Laxness governed by rules (overriding regularities)

(a)	(b)	(c)	(d)
cemetery	tonic	vacuum	vicious
poverty	solid	continue	myriad
ability	blemish	menu	abolition
opera	cherish	value	trivial
animal	tenet	manual	
parody	credit	gradual	
	gravel		

All the examples of (22) above have their stressed vowel-letter in a free graphic position and yet they are all pronounced with their regular lax value. The reason for this is that the effect of the free position rule is invalidated by stronger regularities. In column (a) Trisyllabic Laxness is in force, in column (b) the laxing endings *-ic, -id, -ish, -et, -it, -el* cause laxness in the preceding syllable, in column (c) the vowel letter U in a free graphic position in the next syllable has a laxing effect on the preceding stressed vowel-letter, while in column (d) iCiV laxing takes place. These regularities have all been introduced already in connection with Vowel-Shift in section 3 above, and therefore we will not discuss the details here again.

It is worthy of note that the most problematic environment for letter-to-sound correspondences is the penultimate syllable where we find a lot of examples with a graphically free vowel pronounced lax. Unfortunately these examples must be memorized one by one. We have listed a couple of such words in (23) below to illustrate the point.

Table 23. Irregularities to the free position rule in the penult

melon, lemon, canon, salad, manor, foreign, liver, linen, atom, prison, petrol, body, study, proper, florist, metro, baron, devil, Devon, dragon, leper, baton, basil, Adam, copy, river, pity, eleven, never, Alex, etc.

A final note is due here in connection with column (b) in (20) above. The examples occurring in column (b) of table (20) are connected to the general rule of prevocalic tenseness in English. This rule prescribes that

prevocalic vowels in English must be tense. In other words, lax vowels are excluded from a prevocalic position, and hence they may not appear as a first member of a hiatus. This is illustrated in (24) below where syllable boundaries are indicated with a dot.

Table 24. Prevocalic Tenseness

neon /ˈniː.ən/, fluid /ˈfluː.ɪd/, hiatus /haɪ.ˈeɪtəs/, poem /ˈpəʊ.əm/, flying /ˈflaɪ.ɪŋ/

4. Conclusion

There are a number of regularities both phonological and orthographic to support the viability of the view that the tense-lax division of English vowels is functional rather than phonetic. Trisyllabic Laxness, Pre-r Breaking, Pre-R Broadening, Vowel-Shift, Letter-to-sound correspondences, Prevocalic Tenseness all suggest that the terms *tense* and *lax* must be preserved in the technical vocabulary characterising the English vowel system. More importantly, the tense-lax dichotomy does not coincide with the long short division, and therefore we do not multiply our theoretical entities beyond necessity.

References

Csides, Cs. in prep. *English Phonetics and Phonology: Theory and Practice.* ms

Durand, J. 1990. *Generative and Non-linear Phonology*. London: Longman.

Giegerich, H. J. 1992. *English Phonology: An Introduction*. Cambridge: CUP.

Harris, J. 1994. *English Sound Structure*. Oxford, UK/Cambridge, USA: Blackwell.

Katamba, F. 1993. *An Introduction to Phonology*. London/New York: Longman.

Kreidler, C. 1989. *The Pronunciation of English: A Course Book in Phonology*. Oxford, UK/Cambridge, USA: Blackwell.

Nádasdy, Á. 2006. *Background to English Pronunciation*. Budapest: Nemzeti Tankönyvkiadó.

Wells, J. C. 1982. *Accents of English 1: An Introduction*. Cambridge: CUP.

CONTRIBUTORS

Patricia Ashby is Emeritus Fellow to the University of Westminster and National Teaching Fellow of the UK Higher Education Academy. She holds an MA and PhD in phonetics from University College London and a BA in English from the University of Lancaster. She has taught phonetics and phonology for over thirty years in countries all over the world, including Belgium, Poland, India, Germany and Japan. Her publications include *Speech Sounds* (1995, 2005) and *Understanding Phonetics* (2011). She has an international reputation in the field of phonetics pedagogy. Other research interests include English intonation. Dr Ashby is Examinations Secretary and a Member of Council of the International Phonetic Association, she is Director of the IPA Examination Strand on the long-running UCL Summer Course in English Phonetics and a co-founder and organizer of the biennial Phonetics Teaching and Learning Conference.

Andrej Bjelaković is Teaching Assistant of English Linguistics at the Department of English, University of Belgrade (Serbia), from which he graduated in 2010, and where he received his MA in 2011. The title of his MA thesis was *Changes In RP During the Second Half of the 20th Century*. He is currently a doctoral student in linguistics at the Faculty of Philology, University of Belgrade, and teaches Pronunciation within the Contemporary English G1/G2 course. He attended the Summer Course in English Phonetics at University College London in 2007. His areas of interest include phonetics and phonology, variationist sociolinguistics, dialectology, and history of English.

Tsvetanka Chernogorova is Senior Assistant Professor of English Phonetics and English Language in the Department of English and American Studies, Faculty of Classical and Modern Philology, University of Sofia "St Kliment Ohridski", Bulgaria. Some of the courses that she is currently teaching include English Phonetics, National Varieties of English, as well as an MA-level course in English Phonetics. In 2011 she attended the UCL Summer Course in English Phonetics (EFL Strand). She has taken part in teacher-training seminars for English teachers in Bulgaria, participated in a number of conferences, and published a number of research papers in book collections. In 2013 she has participated in the

Erasmus Life Long Learning Programme and delivered a course of lectures on National Varieties of English at Saarland University, Saarbrucken, Germany. She is also actively involved in an on-going bilateral project "English for the Global Network" whose purpose is to collect a corpus of academic spoken English. Her primary research interests include experimental phonetics, socio-phonetics, the teaching of English pronunciation, English pronunciation models and accents, L1 transfer, L2 phonological acquisition and perception and speech corpora.

Alan Cruttenden is Emeritus Professor of Phonetics, University of Manchester and Fellow of the Phonetics Laboratory, University of Oxford. He studied in Oxford, Wales and London. His doctoral thesis for the University of Manchester was on the intonation of adverbials in English. He was Head of the Department of Linguistics in Manchester from 1992-95. He edited the *Journal of Child Language* from 1984-90. He has lectured in many countries in Europe and in the U.S., Japan and South America. He has published around fifty articles, mainly on intonation and on child language, in journals such as *Phonetica, Journal of Linguistics, Journal of the International Phonetic Association, Lingua, Journal of Pragmatics, Journal of Child Language,* and *English Language Teaching*. He has authored three books: *Language in Infancy and Childhood* (Manchester University Press, 1979, 1985), *Intonation* (Cambridge University Press, 1986, 1997) and *Gimson's Pronunciation of English* (Hodder, 1994, 2001, 2008).

Csaba Csides is Associate Professor of English Linguistics at the English Linguistics Department of the Institute of English Studies at Károli Gáspár University of the Reformed Church in Hungary, Budapest. He studied linguistics and received his MA degree at the Department of English Linguistics, Eötvös Loránd University, Budapest, and completed his PhD studies at the Doctoral School in Linguistics at Eötvös Loránd University, Budapest. He is the author of *Structural Relations and Government in Phonology: A strict CV Approach* (2008) published by VDM Verlag Dr. Müller, Saarbrücken, Germany. Dr. Csides was a member of a board of pronunciation editors of the *English-Hungarian Comprehensive Dictionary* (1998) published by the Academic Press, Budapest and also of the *English-Hungarian Dictionary* (2000) published by Aquila Press, Budapest. Dr. Csides' research focuses on phonetics and phonology, and he regularly gives talks at various international conferences. Currently he is working on a university textbook entitled *English Phonetics and Phonology: Theory and Practice* to be published soon. Besides phonetics

and phonology, he also teaches morphology, introduction to linguistics and syntax at Károli Gáspár University. Dr Csides was the main organizer of the international conference entitled *Interfaces in English Linguistics* hosted by the Institute of English Studies at Károli Gáspár University of the Reformed Church in Budapest, Hungary in October 2012.

Biljana Čubrović is Associate Professor of English Linguistics in the Department of English, University of Belgrade (Serbia), where she received her MA and PhD (in 1999 and 2004, respectively). She is currently a visiting scholar in the Department of Linguistics of Cornell University, where she is carrying out phonetic research. She is the author of two monographs, *Profiling English Phonetics* (2009, 2011) and *The Phonological Structure of Recent French Loanwords in Contemporary English* (2005). She also authored *A Workbook of English Phonology* (2003, 2005), and co-authored three books of tests entitled *English Entrance Exam Practice – tests with key and explanatory notes*. Dr. Čubrović is Editor-in-Chief of *Philologia*, the professional-scientific journal for the study of language, literature and culture (ISSN 1451-5342) and Associate Editor of *The Linguistics Journal* (ISSN 1718-2298; ISSN Print 1718-2301). She has given invited talks in Japan, UK and Serbia, and participated in a number of international conferences. In August 2006, she was awarded a Certificate of Proficiency in the Phonetics of English by IPA. The focus of Dr. Čubrović's work and publications is English phonetics and phonology, as well as linguistic and cultural interplays between English and Serbian, language contacts, and EFL testing. She is a founding member of an international circle of scholars, *English Scholars Beyond Borders* (2013).

Snezhina Dimitrova is Associate Professor of English Phonetics and Phonology in the Department of English and American Studies, the Faculty of Classical and Modern Philology, Sofia University "St Kliment Ohridski", Bulgaria, where she also received her MA and PhD. Some of the courses that she teaches include English Phonology, Varieties of Spoken English, The Languages and Cultures of Scotland, as well as a number of MA-level and teacher retraining courses in phonetics, phonology and the pronunciation of English. Snezhina Dimitrova spent the academic year 2005/2006 working as a Research Associate in the Department of Linguistics and English Language, the University of Edinburgh. She is currently a Fulbright Visiting Scholar in the Department of Linguistics, UCLA. Her primary research interests include experimental phonetics and socio-phonetics, English and Bulgarian phonology, the

prosody of English and Bulgarian, the teaching of English pronunciation, and speech corpora collection and annotation. Dr. Dimitrova has presented at international conferences, participated in a number of bilateral and multilateral research projects, and has published extensively both at home and abroad. She is the author of the book *English Pronunciation for Bulgarians* (2003).

Vladimir Phillipov holds the position of a Senior Lecturer in English Phonetics and Phonology at the Department of English and American Studies at the University of Sofia "St. Kliment Ohridski". He read English at the University of Sofia (1980-1981) and the University of Malta (1982-1987). He was a visiting lector in Bulgarian Language, Literature and Culture at the Universities of Leeds and Sheffield (1992-1995), a HESP scholar at the Research Institute for Linguistics, Budapest, (1996-1999), an Erasmus lecturer in 2002 at the University of Chanakkale, Turkey, and a visiting lecturer in General Linguistics at SUNY, Albany, NY (winter semester, 2006). He participated in various international fora in phonetics: Barcelona, 1995; San Francisco, 1999; Budapest, 1999; Belgrade, 2010; Belgrade, 2012. His scholarly work focuses on intonation from a typological point of view and related issues from general linguistics.

Ken-Ichi Kadooka is Professor of EFL at the School of Business Administration, Ryukoku University (Kyoto, Japan). He received his MA degree at Kobe University in 1990, and finished the doctoral course work in 1995. He published two linguistic books written in Japanese: one on phonology and morphology of Japanese onomatopoeia and the other is a contrastive study of the function and meaning of the intonation systems. Other articles on English phonetics include *Patterns of Clause Intonation in English* in *Ta(l)king English Phonetics across Frontiers* and *Punch Line Paratone in English* in *Exploring English Phonetics*. His current research interest is the Paratone effects in the punch line of jokes with the framework of Systemic Functional Grammar, and former concern was the vowel systems of English, Japanese and Mandarin Chinese. He wrote more than 80 published articles in English and Japanese, presented papers in English in more than 30 international conferences. He also edited 13 English textbooks for Japanese university students, focusing on grammar and pronunciation of English. He is one of the directors of Japan Association of Systemic Functional Linguistics.

Brian Mott is a lecturer in Linguistics in the English Department of the University of Barcelona, where he teaches Phonetics and Phonology,

Semantics and Translation. He was Coordinator of the Linguistics Section of the English Department from 1984 to 1990 and from 2007 to 2010, and a teacher in the independent University Language School ("Escola d'Idiomes Moderns") for thirty years (English coordinator 1976-1980). From 2005 to 2010 he tutored on the Summer Course in English Phonetics at University College London directed by Professor John Wells. He has an MA in Spanish Studies (Aberdeen, 1969) and a PhD in Spanish Dialectology (Barcelona, 1978). Apart from Spanish and Catalan, which he speaks fluently, he also has a knowledge of French, German and Portuguese, and is currently concentrating on Romanian and Serbian. His PhD is a study of the speech of a Spanish Pyrenean village bordering on France in the Province of Huesca, Aragon, and at present he is researching into the intonation of Aragonese varieties. He has also studied the Mirandese dialect spoken in North East Portugal. He has published about twenty books and many articles on all the above areas of specialization. His main extra-academic activity is music. He is a jazz enthusiast and has played bass and sung in various bands.

Yulia Nenasheva is Associate Professor of English Linguistics in the Department of Languages and Interpretation, Magnitogorsk State University (Russia), where she graduated in 1998 summa cum laude. She received her PhD (kandidat nauk) in Nizhniy Novgorod Linguistic University in 2007. She is currently carrying out research in linguistics and supervising undergraduate phonetic research in Magnitogorsk State University. The focus of Dr. Nenasheva's work and publications is English phonetics and phonology, as well as linguistic and cultural correlation between English and Russian, non-verbal behaviour in different cultures, speech patterns. Dr. Nenasheva is the author of the tutorial *Theoretical Phonetics of the English Language (2008, 2009)*, a workbook on English phonetics (2009). She also co-authored two books in the series *Step Forward to Better English* (2012). She has participated in a number of international conferences in Belgrade, Serbia (2012), Budapest, Hungary (2012), and Sydney, Australia (as undergraduate participant supervisor, 2012).

Tatjana Paunović is Associate Professor of English Language and Linguistics in the English Department of the Faculty of Philosophy, University of Niš, Serbia, where she teaches undergraduate and MA courses in linguistic phonetics, acoustics, pronunciation in EFL teaching, and intercultural communicative competence. Her published work includes *English phonetics and phonology for Serbian EFL students* (2007, in

English), and *Phonetics and/ or Phonology? A Critical Review of the 20th century Phonological Theories* (2003, in Serbian). She has presented at international conferences and participated in research projects, currently in the project *Languages and cultures across time and space* (Serbian Ministry of Education, Science and Technological Development). Her research interests include linguistic phonetics, applied linguistics, sociolinguistics, particularly language attitude research, and intercultural communication in the context of EFL teaching.

Aleksandar Pejčić is an English language and linguistics graduate of the Faculty of Philosophy, University of Niš (Serbia), where he received his MA in 2012, for the thesis *What do We Believe? Prosodic Correlates of Persuasive Speech in Serbian and English Political Discourse*. He is currently a PhD student in the Philology program at the Faculty of Philosophy, University of Novi Sad, and a teaching assistant in undergraduate Phonetics and Phonology courses at the Faculty of Philosophy, University of Niš. His research and academic interests include English Phonetics, particularly the acoustic analysis of speech, as well as speech perception, discourse analysis, and contrastive analysis. He has taken part in a published student translation project and has presented papers at international conferences (BIMEP 2012, Belgrade; ELALT 2, Novi Sad).

Oksana Pervezentseva is Professor of English Phonetics at the Faculty of Foreign Languages, Moscow State Pedagogical University (Russia), where she received her PhD and Academic Rank of Associate Professor (in 1997 and 2009, respectively). She is currently a lecturer at the Department of English Phonetics, Moscow State Pedagogical University, where she is carrying out her doctoral phonetic research. She is the author of the monograph, *Building a Prosodic Model of the Speech Act Structural Organisation (in English-Russian Interference Environment)* (2011). She also authored *Theoretical Phonetics of the English Language* (2006) and co-authored *Theoretical Phonetics of English: Practicum (2011)*. She has been a visiting professor in Southern Arkansas University, and participated in a number of international conferences. The focus of Prof. Pervezentseva's work and publications is prosody in discourse and pragmaphonetics, as well as language contacts and EFL teaching.

Stefano Quaino is an Italian phonetician, currently working at Alpen-Adria University in Klagenfurt (Austria), where he received his M.A. and PhD (2008 and 2011, respectively). His main fields of interest are

phonetics and prosody; during his stay as visiting researcher at Swansea University, he worked at the Archive of Welsh English under the supervision of Dr. Robert Penhallurick. He has been invited to several international conferences and given guest lectures in Austria and Slovenia. Dr. Quaino's other interests are dialectology, World Englishes and sociolinguistics.

Rastislav Šuštaršič is Full Professor of the English Language at the Department of English, Faculty of Arts, University of Ljubljana (Slovenia), where he received his MA in 1989 (*English Loan-Words in Slovene – Phonological Adaptation*) and PhD in 1993 (*A Contrastive Analysis of English and Slovene Sentence Intonation*). He is author of a monograph *English-Slovene Contrastive Phonetic and Phonemic Analysis and its Application in Teaching English Phonetics and Phonology (*2005) and co-author of a course book *Present-Day English Pronunciation – A Guide for Slovene Students* (2002). He has participated in a number of international phonetics and applied linguistics conferences around the world, including two International Congresses of Phonetic Sciences - in San Francisco, US (1999) and Barcelona, Spain (2003). In the year 2000, he was a Research Associate at the University of Manchester, working in the field of acoustic phonetics. He teaches English Phonetics and Phonology, English Accents and Dialects, English Semantics and Pragmatics and English-Slovene Translation. He has also taught general linguistics and general phonology. The focus of his work and publications is contrastive English-Slovene phonetics and phonology and various aspects of teaching of English phonetics at university level. He chaired the English Language Section of the English Department in 1997-2006 and was Head of the English Department in 2006/07-2009/10.

Isao Ueda is Professor of Linguistics in the Department of Language and Information Science, Graduate School of Language and Culture, Osaka University (Japan), and Adjunct Professor at the United Graduate School of Child Development of the same institution. He is also a visiting researcher at the Special Education Center, Fukuoka University of Education (Japan). He was a Fulbright exchange scholar at Indiana University (USA) from 1997 to 1998. His research interests range from acquisition of first and second language phonology and functional speech disorders of children to sociophonetics and forensic phonetics. He is author (and co-author) of numerous articles including "Mora augmentation processes in Japanese," *Journal of Japanese Linguistics* 18, (2002); "The developmental path in the acquisition of the Japanese liquid,"

Developmental Path in Phonological Acquisition: Special Issues of Leiden Papers in Linguistics 2, (2005); "Some formal and functional typological properties of developing phonologies," *Gengo Kenkyu (Journal of the Linguistic Society of Japan)* 127, (2005); "Some aspects of English coda acquisition by Japanese learners," *Proceedings of the Third Seoul International Conference of Phonology*, (2005); "Morphological vs. phonological mora augmentation," *Lexicon Forum* 2, (2006); and "The interface between phonology, pragmatics and syntax in nuclear stress misplacement," *Proceedings of the 2009 Mind/Context Divide Workshop*, Cascadilla, (2010). He is currently president of the Phonological Society of Japan.

Jelena Vujić is Associate Professor at the Faculty of Philology, Belgrade University (Serbia). She teaches courses in Descriptive grammar of English I-IV. She is a well-published author who has participated and presented her papers at a number of international conferences and congresses world-wide. Her scientific interests include word-formation patterns in English and Serbian, features of loanwords, aspects of inflection, sociolinguistic aspects of the language of the diasporas, gender-markedness issues, etc. She is the author of the following three books: *Osnovi morfologije engleskog jezika* (2006), *Describing English through Theory and Practice I* (2011) and *Describing English through Theory and Practice II* (2012). She has also supervised a number of PhD and master theses. Dr. Vujić is a member of the following professional associations: International Linguistic Association, North-American Association for Serbian Studies, *Philologia*, Serbian Society for Applied Linguistics and Serbian Society for the Study of English. In addition, she serves as a member of the editorial board of several language journals, including *ESP Today*.

INDEX

A

accent, 18, 21, 24, 27, 41, 46, 47, 48, 49, 51, 52, 53, 56, 57, 58, 61, 65, 67, 68, 90, 93, 94, 152, 191, 195, 197, 204, 205, 208, 211, 215, 216, 217, 218, 219, 221, 222, 224, 225, 227, 228, 229, 230, 231, 232, 233, 234, 235
acquisition, 13, 14, 20, 21, 22, 23, 25, 135, 160, 228, 284, 289
affix, 251, 252, 254, 255, 256
alignment, 46, 47, 48, 50, 54, 62, 64, 66, 193, 212
articulation, 3, 5, 7, 9, 10, 11, 13, 107, 109, 118, 122, 185, 186, 189, 190, 233
Ashby, 165, 168, 181, 247, 283
ASR. *See* Automatic Speech Recognition (ASR) software
Automatic Speech Recognition (ASR) software, 100

B

back, 5, 6, 7, 9, 14, 15, 16, 20, 21, 22, 29, 30, 32, 33, 39, 92, 115, 117, 176, 184, 218, 231, 242, 255, 262
Best, 14, 24
Birmingham School, 117
British English, 9, 25, 45, 123, 130, 149, 195, 216, 263, 268
Bulgarian, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 115, 121, 122, 215, 216, 218, 219, 224, 225, 285, 286

C

Celtic English, 45
CiV laxing, 261
clear [l], 10
cluster analysis, 154, 155
Communicative Dynamism, 86
constituent order, 86, 96, 97
contour, 46, 47, 49, 50, 51, 52, 53, 54, 55, 56, 58, 61, 62, 63, 64, 67, 68, 72, 75, 76, 77, 79, 80, 83, 109, 193
contrastive, 13, 24, 84, 94, 95, 97, 183, 186, 189, 286, 288, 289
Critical Period Hypothesis, 228
Croatian, 116, 210, 211, 212
Cruttenden, 3, 10, 11, 15, 18, 20, 24, 45, 68, 86, 97, 123, 129, 183, 190, 217, 225, 284
Crystal, 27, 36, 42, 134, 138, 149
Czech, 86

D

dark [ɬ], 10
declaratives, 45, 56, 205
devoicing, 109, 216, 230, 232, 234

diphthongs, 29, 30, 57, 59, 62, 68, 169, 218, 229, 230, 231, 262, 263, 264, 268
discourse topic, 191, 192, 193, 194, 195, 196, 197, 199, 200, 202, 203, 204, 206, 209
disfluencies, 101, 105, 107, 108
downstep, 193, 198, 200, 205
Dynamic MRI. *See* MRI

E

Early Modern English (EME), 27, 28
EFL, 191, 192, 194, 195, 196, 200, 201, 202, 203, 205, 206, 207, 208, 209, 230, 231, 234, 236, 283, 285, 286, 287, 288
elicitation technique, 102, 103
EME. *See* Early Modern English
emotional speech, 101, 111, 157

F

F1 and F2, 13, 17, 19, 22
Firth, 117
Flege, 14, 19, 21, 23, 24, 25
flipped classroom, 166, 169, 172, 179, 180
flipping, 166, 169, 170, 171, 172, 173, 175, 176, 177, 178, 179, 180
Flipping, 166, 169, 171
foreign accent, 227, 233, 234, 235
formant, 17, 18, 19, 21, 22, 23

formant frequencies, 17, 18
fricatives, 33, 53, 185, 186, 189, 218, 230, 232
front, 5, 6, 7, 9, 14, 15, 16, 17, 18, 21, 22, 29, 37, 79, 89, 91, 168, 170, 189, 190, 218, 230
fronting, 10, 20, 21
Functional Sentence Perspective, 86
fundamental frequency, 106, 107, 153, 156, 158

G

General American, 28, 215, 217
generativist tradition, 115, 119
Gimson, 3, 4, 9, 10, 11, 24, 45, 68, 186, 190, 225, 284
Gimson's Pronunciation of English, 3, 4, 11, 24, 190, 225, 284
Gwynedd English, 45, 47

H

Halliday, 76, 84, 117, 118, 120, 123, 130, 138, 149
hard palate, 5, 7, 9
high vowels, 13, 14, 15, 16, 244

I

intensity, 48, 50, 51, 54, 59, 61, 62, 67, 68, 105, 106, 107, 108, 109, 116, 122, 144, 145, 153, 156, 158, 191, 193, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 262
interference, 135, 151, 153, 156, 157, 159, 160, 219, 225, 236, 253
intonation, 45, 47, 48, 71, 72, 73, 75, 81, 84, 85, 87, 89, 95, 96, 97, 108, 109, 110, 115, 116, 117, 118, 119, 120, 122, 123, 124, 125, 126, 127, 128, 133, 134, 135, 137, 138, 139, 147, 148,

149, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 171, 191, 192, 194, 195, 197, 210, 211, 212, 219, 283, 284, 286, 287

J

Japanese, 84, 184, 212, 237, 239, 240, 241, 242, 243, 244, 247, 248, 286, 289

K

Khan Academy, 165, 166, 169, 182

L

L1 transfer, 13, 16, 18, 21, 23, 24, 192, 208, 209, 284
Ladd, 47, 68, 86, 97, 120, 130, 135, 150
lax, 15, 35, 36, 123, 261, 262, 263, 264, 265, 267, 269, 270, 271, 272, 273, 274, 275, 276, 278, 280, 281
Laxing by ending, 261
Laxing by free U, 261
length, 3, 13, 17, 24, 57, 58, 72, 105, 121, 122, 138, 139, 169, 237, 246, 261
level ordering, 251, 254, 255, 257
Lexical Phonology and Morphology model, 251

M

Magnetic Resonance Imaging. *See* MRI
Middle English, 28, 29, 32, 41
minimal pairs, 183, 184, 185, 186, 187
morphology, 46, 251, 253, 254, 255, 258, 259, 285, 286
Most Prominent Semantic Element. *See* MPSE
MPSE, 48, 52, 56, 64, 65
MRI, 3, 4, 6, 9, 10
MRI videos. *See* MRI

N

normalization, 105, 106
nuclear accent, 191, 197
nuclear stress, 85, 88, 93, 94, 95, 96, 110, 290
nucleus, 48, 85, 86, 87, 88, 89, 90, 91, 94, 95, 96, 98, 122, 124, 125, 126, 154, 171, 198, 201, 243, 244, 245, 246

O

Optimality Theory, 251, 256, 257, 258
Ordering of affixes, 251
Original Pronunciation (OP), 27

P

paratone, 71
pauses, 73, 101, 106, 107, 108, 109, 110, 112, 193, 197, 206
Perceptual Assimilation Model – PAM, 14
persuasive Speech, 100, 288
phonetic similarity, 239, 240, 242, 247
Phonetics Laboratory, University of Oxford, 3, 11, 284
pitch accents, 46, 48, 49, 53, 54, 55, 61, 64, 67, 109, 138

pitch range, 105, 191, 193, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 208, 209, 212
political speech, 99, 109, 112
postlexical, 115, 120, 257
pragmatic, 85, 87, 94, 115, 120, 151, 153, 154, 156, 159, 212, 256
Prague School, 86
pre-/r/ breaking, 37
Pre-cluster laxing, 261
Pre-R Broadening, 268, 269, 270, 277, 281
pre-schwa laxing, 37
present-day English (PDE), 27, 28
pronunciation, 18, 19, 20, 23, 24, 25, 27, 28, 30, 31, 41, 47, 49, 57, 61, 93, 101, 105, 107, 108, 111, 135, 136, 153, 183, 184, 186, 189, 190, 215, 216, 217, 218, 219, 220, 221, 223, 224, 225, 231, 234, 261, 271, 278, 279, 284, 285, 286, 287
prosodic markedness, 154
punch line, 71, 73, 74, 76, 77, 78, 79, 80, 82, 83, 286
Punch Line Paratone, 71, 84, 286

R

regional accents, 27
repairs, 99, 101, 107, 108
Rhondda Valley English (RVE), 46
rhythm, 150, 171, 216, 219, 229
rising tones, 45, 48
Romanian, 86, 287
RP, 4, 10, 11, 17, 25, 29, 33, 45, 215, 216, 217, 218, 222, 224, 226, 231, 263, 264, 268, 283
Russian, 121, 122, 123, 124, 126, 152, 156, 159, 161, 227, 229, 230, 231, 232, 233, 234, 235, 287, 288
RVE. *See* Rhondda Valley English

S

SAWD, 46
segmental, 47, 69, 100, 101, 109, 116, 120, 193, 215, 216, 220, 227, 229, 237, 241, 245
Serbian, 20, 25, 30, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 99, 100, 103, 104, 105, 108, 109, 116, 191, 192, 193, 194, 195, 196, 202, 203, 205, 206, 207, 208, 209, 212, 213, 227, 229, 230, 234, 235, 236, 285, 287, 288, 290
Shakespeare, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 171
Slavic languages, 15
Slavonic, 86, 121, 126
Slovene, 87, 90, 91, 93, 97, 183, 184, 185, 189, 190, 212, 289
soft palate, 5, 6, 9
South Slavic, 227
Spanish, 85, 86, 97, 150, 161, 231, 287
speaking styles, 100, 102, 104
Speech Learning Model – SLM, 14
spontaneous speech, 99, 100, 101, 102, 103, 105, 107, 108, 110, 111, 112, 138, 196
stereotype, 34, 134, 135, 136, 137, 148
stress, 40, 41, 48, 85, 86, 87, 88, 89, 95, 97, 119, 152, 153, 159, 171, 211, 215, 221, 229, 252, 253, 254, 255, 257

suprasegmental, 137, 211, 215, 229, 232
Survey of Anglo-Welsh Dialects. *See* SAWD

T

teeth ridge, 5, 6, 7
Tench, 45, 48, 69, 71, 83, 84
tense, 15, 37, 118, 123, 230, 261, 262, 263, 264, 265, 266, 267, 269, 272, 273, 274, 275, 276, 277, 278, 279, 281
TG, 251, 253
ToBI, 46, 47
tongue, 3, 4, 5, 6, 7, 10, 15, 16, 20, 36, 187, 219, 225, 261
tonic stress, 85, 88
tonic syllable, 48, 171
trademark, 237, 238, 239, 240, 241, 243, 247, 248
trademarks, 237, 239, 240, 241, 242, 243, 245, 247
transliterated, 237, 242, 243, 247
Trisyllabic Laxness, 261, 265, 266, 267, 269, 273, 280, 281

U

unrounded, 15, 16, 32, 34, 36, 37, 190, 231, 261
uvula, 5, 9

V

variation, 9, 10, 35, 73, 96, 103, 104, 139, 235
velar nasal, 230, 234
velars, 9
Vowel Shift, 261, 272

W

Wells, 28, 33, 34, 37, 39, 42, 85, 86, 90, 98, 147, 150, 217, 218, 219, 226, 264, 268, 282, 287
Welsh English, 45, 46, 48, 50, 67, 68, 289
Wennerstrom, 71, 84, 107, 113
word-formation, 251, 252, 253, 259, 267, 290

X

X-ray, 3, 4

Z

zone conception, 151, 154, 159