Exploring English Phonetics
Exploring English Phonetics

Edited by

Tatjana Paunović and Biljana Ćubrović
TABLE OF CONTENTS

Introduction ........................................................................................................ v

Part I: Phoneme and Beyond

Vowel Frequencies in Traditional Cockney and Popular London Speech ................................................................. 3
Brian Mott, Spain

Some Controversies about /v/ in Serbian, Transfer in English and Other Related Issues ................................................. 13
Maja Marković and Bojana Jakovljević, Serbia

VOT Transfer in the Production of English Stops by Serbian Native Speakers................................................................. 31
Bojana Jakovljević, Serbia

The Evolution of a Phonetic Phenomenon: The Case of Voice Onset Time in Serbian Intermediate EFL Learners ......................... 47
Biljana Ćubrović, Serbia

Pronunciation in Connected Speech: A Survey of Weak Forms in a Spoken Corpus of American English................................. 57
Takehiko Makino, Japan

Tonic Misplacement by Japanese Learners of English ........................................ 73
Isao Ueda and Hiroko Saito, Japan

Punch Line Paratone in English ................................................................... 85
Ken-Ichi Kadooka, Japan

The Sign Character of Intonation ................................................................ 99
Vladimir Phillipov, Bulgaria
# Part II: Applied Phonetics and Beyond

The Role of Modernized Prescriptivism in Teaching Pronunciation to EFL University Students ................................................. 113  
*Tvtiko Prćić, Serbia*

Hearing the Difference: An Innovative Approach to the Teaching of Pronunciation ................................................................. 125  
*Ingrid Pfandl-Buchegger, Isabel Landsiedler and Milena Insam, Austria*

Qualitative Methods in Phonetic Research: A *contradictio in adjecto*? .................................................................................. 145  
*Tatjana Paunović, Serbia*

Research Strategies in L2 Phonological Fieldwork Investigation and Significance and/or Reliability of Results................................. 163  
*Klementina Jurančič Petek, Slovenia*

Using Web Technologies in L2 Phonological Research: Methodological Issues and Implications................................................. 179  
*Anastazija Kirkova-Naskova and Dimitar Trajanov, Macedonia*

Learning from Students’ Errors: English Phonetics Theory Exam....... 197  
*Rastislav Šuštaršič, Slovenia*

English Pronunciation Models and Tertiary-level Students: A Bulgarian Perspective ................................................................. 207  
*Snezha Dimitrova and Tsvetanka Chernogorova, Bulgaria*

English Pronunciation Norms and the Case of Russian English ........ 225  
*Galina M. Vishnevskaya, Russia*

Contributors............................................................................................. 241

Index........................................................................................................ 251
INTRODUCTION

*Exploring English Phonetics* is a volume comprising sixteen contributions by authors from different linguistic and academic backgrounds and from eight different countries: Austria, Bulgaria, Japan, Macedonia, Russia, Serbia, Slovenia, and Spain. Conceived as a meeting point of diverse perspectives, approaches and interests of scholars working in the field of English Phonetics worldwide, this volume focuses on the topics of language variety, mutual language influences, and issues in researching, studying, and teaching English with speakers of other languages.

Authors raise a number of questions, interesting from the point of view of either phonetic research or phonetic training and EFL teaching. These questions span a wide range of phonetic topics, from the nature of vowels and consonants in particular English varieties, and the phenomena of connected speech or the nature of intonation, via issues in the methodology of phonetic research, to problems encountered by speakers of other languages trying to acquire English pronunciation, and attitudes to different native and non-native varieties of English. Still, despite such a broad variety of topics, the volume does not lack unity. Whatever their focus, most chapters deal with English spoken and learned by speakers of other languages, thus highlighting both the current status of English as the language of world-wide communication, and the international orientation of this volume.

***

The eight papers grouped in the first section of the book, *Phoneme and beyond*, focus on the segmental and prosodic properties of English. In chapter one, Brian Mott investigates the formant frequencies of vowels identified in male London speakers, whose speech the author defines as representing "the range of articulation covered by the continuum running from the basilectal extreme, Traditional Cockney, to Popular London Speech". The study involved a reading task, word tokens illustrating vowels in the /h-d/ context, and a comparison of the speakers' average formant frequencies to the values offered in literature on British English. The author points out the areas in which some observable differences could be identified, namely, the close vowels of *heed* and *who'd*. 
In their discussion of the controversies concerning the phonetic properties and the phonological description of the consonant /v/, found in several languages, Maja Marković and Bojana Jakovljević review the main points of dispute raised in literature about the ‘hybrid’ nature of this consonant, which can be defined as either a labiodental fricative or a labiodental approximant. The focus on the acoustic properties of this consonant in English and Serbian L1 speakers, as well as the evidence of transfer in Serbian L2 English speech.

Voice Onset Time (VOT) is another topic widely discussed in phonetic work today, and investigated in the next two chapters of this volume. The study reported by Bojana Jakovljević focused on the VOT found in British English and Serbian word-initial stops at the beginning of a stressed syllable, seeking to identify possible transfer in the speech of experienced Serbian L2 English learners. The study involved female speakers, and reading tasks with word lists illustrating stops in matching vowel contexts in Serbian and English. The author points out that although the participants were L2 English speakers with many years of experience and pronunciation training, the findings revealed obvious L1 transfer in the production of English word-initial, particularly voiced, stop consonants.

The next chapter, by Biljana Ćubrović, investigates the Voice Onset Time phenomenon with intermediate Serbian EFL speakers, but focuses more narrowly on the production and acquisition of aspirated voiceless stops /p t k/. The study involved male and female speakers, a reading task with selected monosyllabic English words, and the comparison of the acoustic properties of the participants' consonants with their matching Serbian counterparts. The findings of this study suggest that aspiration is acquired gradually and that it is a phonetic phenomenon worth insisting upon in an EFL classroom.

The chapter by Takehiko Makino moves on to the level of connected speech to discuss the use of weak forms in colloquial American speech. The author points out that the descriptions of weak forms found in literature focus inappropriately on relatively formal styles of speech, while less formal and colloquial speech reveals many more examples of weakening that should be taken into account. The study included examples found in the Buckeye Corpus of Conversational Speech. The author suggests further investigation of a range of less formal styles.

Isao Ueda and Hiroko Saito focus on problems of tonic placement with Japanese English speakers. The authors set off from the observation highlighted by previous research that Japanese-accented English is often characterized by tonic (nuclear stress) misplacement. The study reported in this chapter involved fifteen university students, English majors, and a
longitudinal aspect, i.e. a pre-test and a post-test repeated with the same participants after a year of studying English in various courses, including those focusing particularly on the listening and speaking skills. The study investigated the participants' production of tonic stress, as well as their phonological awareness of tonic stress placement in English. The findings showed that the acquisition process with Japanese EFL learners involves a transition from one type of tonic placement to another. The authors conclude that for some learners formal instruction of tonic placement may be more important and effective than practical training.

The chapter by Ken-Ichi Kadooka introduces the topic of paratone, a phonological category functioning in spoken language as a counterpart of paragraph structure in written language. The author focuses on a proposed subtype of paratone, which he terms punch line paratone, found in the genre of jokes, in the last line that carries the main point, as a rhetorical effect used to emphasize the main point of the joke. The author analyzes examples of English jokes and compares them with Japanese Kobanashi stories. The analysis suggests that both types of stories share some characteristics, including the punch line tone pattern, and a brief pause inserted before the punch line.

From a more theoretical perspective, Vladimir Phillipov discusses the sign nature of intonation, viewing it through the prism of different theoretical approaches, as a 'co-sign' (Bulgarian пръ-знак, German Anzeichen), or ‘an indexical sign’. The author stresses the fact that intonation permeates all linguistic levels, drawing together grammatical, semantic and pragmatic functions.

The second part of this volume, Applied Phonetics and beyond focuses even more narrowly on teaching and pronunciation acquisition with specific EFL groups of speakers, but also on some issues of phonetic research methodology, particularly important from the EFL perspective.

Tvrtko Prčić discusses the notion of 'modernized prescriptivism' in EFL pronunciation teaching, that is, in the pronunciation training within the language the author refers to as "the nativized foreign language (ENFL)". The author sets off from the discussion of some theoretical and methodological aspects of descriptivism and prescriptivism in previous studies, moving on to describe the main principles of the proposed modernized prescriptivism, and to offer some specific methodological and practical hints on how it can be applied in pronunciation training. The author also points out the necessity of introducing an integrated approach to teaching language and linguistics, and the use of what he terms ‘usage-enriched descriptivism’, which combines descriptivism with elements of modernized prescriptivism.
In the next chapter, Ingrid Pfandl-Buchegger, Milena Insam, and Isabel Landsiedler describe an innovative L2 teaching project, titled *FauvoT*, implemented at the University of Graz, Austria, which aims to accelerate L2 learning through focused listening. The authors offer a description of the materials and students' activities in language laboratory and at home, as well as an analysis of students' results and the evaluation of their performance. Since the study showed that the participants made an obvious improvement in pronunciation, it is suggested that the use of electronically modified tapes and concentrated listening, providing students with a more focused perception of foreign language sounds, is an efficient tool for L2 phonological acquisition, particularly with adult learners.

Phonetic research methodology is discussed from different angles in the next three chapters. Starting from the observation that most research in the field of phonetics is still conducted within the traditional experimental paradigm, Tatjana Paunović asks whether, and in what ways, phonetic research could benefit from widening the perspective to include some elements of qualitative methodology. The author uses three of her own research studies, based on traditional methodology and techniques and focusing on EFL speakers, to illustrate how including a qualitative perspective and a mixed-method research design could contribute to a better understanding of L2 speech.

Klementina Jurančič-Petek discusses the influence of data-gathering methods on the nature of the results obtained and the reliability of conclusions drawn from data in phonetic and phonological investigations. Starting from the fact that research findings may or may not be influenced by the administration of different types of tests (sentence reading tasks, phrase reading tasks, imitation, free speech), the author discusses the observations of some previous authors concerning factors such as participants' age, attention span, level of proficiency etc., and goes on to examine some of these factors on the example of the pronunciation of English by Slovene learners. The author points out that the type of data gathering procedure in her study indeed resulted in observable differences in the participants' results, and suggests that more studies should involve free or spontaneous speech, in addition to more traditional reading techniques.

The chapter by Anastazija Kirkova-Naskova and Dimitar Trajanov also focuses on research methodology in phonetics. The authors describe a study of L1 Macedonian English learners' pronunciation, that is, the perception of Macedonian English speech by a number of native speakers. The study involved a group of 1st- and 2nd-year students of English at
Skopje University, whose speech was recorded and then evaluated by trained phoneticians, native speakers of English. By using a specially designed web application as a data gathering instrument, it was possible for the researchers to include a great number of trained assessors in the research. The authors draw attention to the fact that new Internet-based technologies make it possible to apply modern and more carefully designed research methods even in unfavourable circumstances. Such technologies also have the advantage of being flexible and thus more widely applicable, when adapted to specific research aims and questions.

Rastislav Šuštaršič focuses on the linguistic (phonetic) education of English language majors at the English Department, University of Ljubljana, Slovenia. The author discusses some problems observed in the oral exam in English phonetics, i.e. the theoretical part which tests students’ knowledge of the English sound system, and the allophonic realizations of vowels and consonants including aspiration, glottalization, voicing and devoicing, vowel duration, assimilation, etc. Since students are required to demonstrate their understanding of certain phonetic notions on the specific examples of particular words or phrases illustrating these notions, the author concludes that students' mistakes and problems in this exam can be used as a valuable guide for teaching, highlighting particularly problematic areas in which students need to be offered additional explanations.

The last two chapters discuss the choice of the model in teaching L2 English pronunciation. Using the example of Bulgarian EFL learners, Snezhina Dimitrova and Tsvetanka Chernogorova start by pointing out the differences between the ELF perspective and the native-speaker model perspective, stating that opting for the latter raises yet another question, that of which native-speaker model to choose. The authors present the results of several surveys they have conducted with tertiary-level English students at Sofia University to investigate their opinions of and attitudes towards different varieties of English. The results are compared to the findings of similar surveys in Bulgaria and other parts of Europe (Poland, Spain). The authors conclude that Bulgarian university students of English still seem to favour the standard British (RP) model of pronunciation.

Focusing on Russian L2 English learners, Galina M. Vishnevskaya states that the choice of the pronunciation model, in the context of such a great variety of pronunciation possibilities in English today, when even the formerly unacceptable 'foreign accents' are recognized as legitimate varieties of English, presents a problem for both teachers and learners. Focusing on some prosodic variables of accented speech, the author describes a study in which the perception of accented speech was
investigated. The assessors were native English speakers, 40 students of University of Boston, USA, and non-native English speakers, 50 students of Ivanovo State University, Russia. The findings showed that native speakers had a more 'tolerant' view of accented speech. Furthermore, the author suggests that a distinction can be made between heavily accented speech, which prompted a very negative reaction, and a slight accent, which provoked positive attitudes. The author, therefore, suggests that a distinction should be made between accent problems that are only phonetically observable, and thus not important in L2 teaching, and those that are communicatively relevant, because they affect the learner's success in communication.

***

Exploring English Phonetics aims to draw attention to issues that can be of interest to both phonetic researchers and applied phonetic practitioners or EFL teachers, and, in some parts, even to a wider audience. Some of the topics dealt with in this volume are among the most widely discussed ones today, from different perspectives, and not only in academic circles. We hope that this volume offers a valuable contribution to this discussion through the different voices of sixteen chapter authors.

The Editors
October, 2011
PART I.

PHONEME AND BEYOND
VOWEL FREQUENCIES IN TRADITIONAL COCKNEY AND POPULAR LONDON SPEECH

BRIAN MOTT

Outline

To calculate the vowel frequencies of speakers whose speech might be described as being somewhere within the range of articulation covered by the continuum running from the basilectal extreme, Traditional Cockney, to Popular London Speech, recordings were made of three men from London, aged 55, 63 and 67 at the time of the exercise, reading the vowels in the context /h-d/ three times each. The averages obtained for the F1 and F2 of each of the vowels were compared with those given for male speakers in Wells (1962) and Deterding (1997). The vowels were found to be similar to those of RP in some cases, like *hid* and *hood*, but not all, e.g. *heed* and *who’d*, which had lower and fronter vowels.

Figure 1. Greater London
1. Introduction

Strictly speaking, Cockney is the basilectal extreme of the popular speech of London, used in an imprecise area north of the River Thames referred to as the East End. The traditional core neighbourhoods of the East End are Bethnal Green, Stepney & Poplar (since 1965 forming the borough of Tower Hamlets), Shoreditch, Hackney, Mile End and Bow, and a little further south, nearer the river, Spitalfields, Whitechapel, Wapping, Limehouse and Millwall. Nowadays, certain areas south of the river (Southwark, Bermondsey and Walworth) are also strongly associated with Cockney speech. However, most of the time, the term “Cockney” is applied loosely to any working-class London accent that deviates noticeably from the standard (RP or SSB, as it is variously called). Accents closer to the standard might be termed Popular London Speech.

In recent years, much has been written on the presumed influence of the speech of London on that of regional varieties of English spoken outside the capital. Speech which shows features associated with London, like T-glottalling, L-Vocalization, HappY Tensing and Yod-Coalescence is sometimes referred to as Estuary English, though the term is seen less in print latterly and may be falling into disuse.

In the present paper, there is no intention to present Traditional Cockney and Popular London Speech (henceforth TC and PLS) as two distinct varieties, but rather as a continuum. Note also that no attention will be paid to more recent innovations in the vowel system of some present-day Inner-London speakers who use monophthongized versions of the FACE and GOAT vowels, which in TC and PLS are traditionally pronounced more like [AI] and [AU], respectively. Traits of this kind will be considered to form part of Multi-cultural London English, which is outside the domain of this study.

Cockney is generally a low-prestige variety, but it also has covert prestige through characters such as Liza Doolittle in G. B. Shaw’s Pygmalion and Sam Weller in Dickens’ Pickwick Papers, who pronounced his [v]’s rather like [w]’s and said such things as “wery good”.

People often associate Cockney with rhyming slang like plates of meat ‘feet’ and trouble and strife ‘life’, but this phenomenon is in fact very marginal and not as common as is believed. Some binomial items of this kind, like butcher’s = butcher’s hook ‘look’, loaf = loaf of bread ‘head’ and china = china plate ‘mate’, have spread into General English. In more recent times, the trend of occasionally spicing one’s language with these rhymes has led to creations based on the names of famous people, like Hank Marvin ‘starving’ and Shania Twain ‘pain, nuisance’.
2. Some generalizations about Cockney monophthongs

The preliminary observations on Cockney vowels that I am going to make are either well known to phoneticians or have been gleaned from the literature.

Figure 2. The (relatively) pure vowels of Cockney

<table>
<thead>
<tr>
<th>RP</th>
<th>Symbols used for Cockney</th>
<th>Usual range of variation in TC &amp; PLS</th>
<th>Sample word</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. /i/</td>
<td>[ɪ]</td>
<td>[i ~ əi]</td>
<td>bee [bɪi]</td>
</tr>
<tr>
<td>2. /ɪ/</td>
<td>[ɪ]</td>
<td>[ɪ ~ ɪ]</td>
<td>bit [bɪʔ]</td>
</tr>
<tr>
<td>3. /ɛ/</td>
<td>[ɛ]</td>
<td>[ɛ]</td>
<td>bed [bɛd]</td>
</tr>
<tr>
<td>4. /æ/</td>
<td>[æ]</td>
<td>[æ ~ ɛ]</td>
<td>mat [mæʔ]</td>
</tr>
<tr>
<td>5. /ɑː/</td>
<td>[ɑː]</td>
<td>[ɑː ~ ɑː]</td>
<td>Margate ['mɑːɡəʔ]</td>
</tr>
<tr>
<td>6. /ɔ/</td>
<td>[ɔ]</td>
<td>[ɔ ~ ɔ]</td>
<td>jot [dɔʔ]</td>
</tr>
<tr>
<td>7. /ɔː/</td>
<td>[ɔː]</td>
<td>[ɔː ~ ʊ ~ ɔ]</td>
<td>yawn [jɔn]</td>
</tr>
<tr>
<td></td>
<td>[ɔː]</td>
<td>[ɔː ~ owɔ]</td>
<td>yourn [jɔrn] ‘yours’</td>
</tr>
<tr>
<td>8. /ʊ/</td>
<td>[ʊ]</td>
<td>[ʊ ~ ʊ]</td>
<td>look [luʔ]</td>
</tr>
<tr>
<td>9. /uː/</td>
<td>[uː]</td>
<td>[uː ~ əʊ]</td>
<td>loopy ['luʊʔi] ‘mad’</td>
</tr>
<tr>
<td>10. /ʌ/</td>
<td>[a]</td>
<td>[a ~ ʊ]</td>
<td>London ['lʌndən]</td>
</tr>
<tr>
<td>11. /ɔː/</td>
<td>[ɔː]</td>
<td>[ɔː ~ ɔ ~ əː]</td>
<td>nurse [nɔs]</td>
</tr>
<tr>
<td>12. /ə/</td>
<td>[ə]</td>
<td>[ə ~ ə]</td>
<td>water ['woːə]</td>
</tr>
</tbody>
</table>

The symbols I use in the second column of figure 2 are those found in Wells (1982: 304), except for the STRUT vowel, for which I prefer open [a] in accordance with its closeness to cardinal vowel 4 in broad accents. The range of variation of the vowels shown in column 3 was constructed from examples and descriptions culled from the literature, notably Wells 1982: 303-321.

As can be seen from figure 2, the short vowels are often similar to those of RP, especially in less broad varieties of Cockney, though the field of dispersion of the allophones may not coincide exactly. For example, the KIT vowel may be more central than in RP, and the TRAP and LOT vowels less open: [ɛ] and [ɔ], respectively (Wells 1982: 305). The pronunciation of the word Saturday, for example, is sometimes ['sɐʔdəri]. In a few words like gawd (god), gone, off and cough, the long vowel [ɔː] instead of [ʊ] may still be heard from the older generation of Cockneys, but it is
recessive and often ridiculed or used in jocular expressions like *Now you've been and gone and done it!* ['nɛː jɔːn 'bɛn ən 'gɔn ən 'dan i?].

Note also that [ʊ] can be more fronted than in RP in some instances, notably the adjective *good* [ɡʊd], and that RP [ʌ] is in general noticeably more open in London speech, sometimes resulting in [a], as in *come* [kʰam], present and past tense in Cockney of the verb *to come*. Schwa is also perceptibly more open in word-final position: *dinner* ['diːnər].

Instead of the open monophthong [e], broad Cockney may occasionally have closer allophones with a palatal off-glide before a voiced consonant: *bedroom* [bɛidrəʊm], *leg* [leiɡ], the beginning of the diphthong being perhaps a little further back than [e] (See O’Connor 1973: 156).

As for the vowels corresponding to RP long vowels, these are often appreciably more diphthongal than in RP. The *FLEECE* vowel tends to close after beginning with a more open tongue position than in RP, which may be as low and centralized as schwa, as in the name *Steve* [ʃtiː] ~ [ʃtəi]. It is usually diphthongal too when word/morpheme-final and unstressed, as opposed to RP [i] (the *happY* vowel), in words like *busy* ['bɪzɪ], and also where older RP has an unstressed *KIT* vowel word-initially, as in *effect* [ɪˈfɛkt], *electric* [ɪˈlektrɪk] and *economy* [ɪˈkɒnəmi].

The equivalent back vowel (the *GOOSE* vowel) is similarly slightly diphthongal, beginning more open and centralized than in RP and gliding to a higher, generally centralized position ([uɻ]). In recent times, this vowel has shown a tendency to become much fronter while retaining some of its rounding: [yː]. This is particularly noticeable when the vowel is preceded by [j], as in *you* [jyː]. Kerswill & Williams (2005) refer to the proliferation of this vowel outside London.

The *PALM* vowel has a fully back allophone ([æː]) considered to be a marker of broad Cockney, while the *THOUGHT* vowel tends to be higher than in RP in closed syllables ([ɔː]) and very often diphthongal, with a glide in the region of [ʰɔʊ ~ əʊ], and centering in open syllables ([ɜː ~ ɔː]), including derivatives ending in a consonant. Thus *board* [boʊd] in phonological opposition to *bored* [bɔːd] < *bore* [bɔː] (this phonemicization is referred to in Wells (1982: 310) as the *THOUGHT Split*. The *NURSE* vowel may be slightly fronted and/or slightly rounded with allophones in the [ɹː ~ əː] range.

### 3. Vowels in hiatus

As this paper is concerned with the monophthongs of London speech, I shall not mention glide insertion after diphthongs ending in the *KIT* and
FOOT vowels. However, it should be mentioned that high monophthongs followed by another vowel may undergo glide insertion, as in freer ['fraɪə], fewer ['fjuərə], piano [piənə] and influence ['ɪnfləns]. Moreover, even if glide insertion is not produced, such words are not subject to compression in TC or PLS.

There are also cases of the definite article followed by a vowel, like the end of the road [ði ˈend ə ˈroʊd], the artful dodger [ði ˈɑːfəd ˈdɒdʒə] and where the hell (have) you been? [ˈweər ˈhi ˈhæv bɪn], or the prepositions to and through plus a vowel: we went to a party [wi ˈwɛnt ə ˈpaɪti]; he drove through a wall [hi ˈdrɑʊf ə ˈwɔl].

Although Cockney prefers to maintain contiguous vowels uncompressed or to use glide insertion as a solution to avoid hiatus or resort to linking/intrusive [r], there are instances of the compression of unstressed syllables not involved in vowel hiatus: I suppose so ['spaʊsəʊ], for instance [fə ˈɪnstənz], perhaps [pəˈrɛps], because [kʰoʊz] ~ [kʰɔz]. Note also the drastic reduction in fast speech of the adverbs actually ['æktʃuəli] > ['æktʃuəli] > ['ækli] and usually ['juəzjuəli] > ['jʊəζli], which do contain vowels in hiatus.

One particularly interesting case of compression involves the deletion of schwa, usually, though not always, representing the indefinite article. Wells (1982: 321) draws attention to this as a neglected phenomenon and observes that it may occur when schwa is preceded by a glottalled [t], as in better have another one ['bɛtər ə vər əˈnəʊr wæn] (my transcription) and about a week ['əbət ə wɛk]. From my own personal experience, I have the example You got a(n) invite, like? [ju əˈɡoʊ əˈɪnvɪt əˈlaɪk] ‘Have you got an invitation?’, in which the indefinite article is preceded by a glottalled [t] and followed by a vowel, but examples like half a(n) hour ['half ə ˈhaʊr] and after (a) hard day’s work ['ɑːftər əˈdaiərd ˈwɜːk] show that the preceding segment in the environment of a deleted indefinite article need not be a glottalled [t]. In fact, what these examples show is that there is a tendency in Cockney to use the indefinite article a, and not an, even before vowels, and that it may then be deleted. There is also a tendency to use the definite article [ðə] rather than [ðəi]. Sue Fox in her unpublished PhD dissertation suggests that the attrition of the allomorphy of both articles is a diffusing innovation from within the ethnic minority community (Britain 2007: 104). However, just how recent the phenomenon is remains a moot point. Dickens seems to have been aware of it, as there is at least one instance in the speech of Mr. Bumble: “...the law is a ass—a idiot” (Oliver Twist, chaper 51).
4. The frequencies of Cockney vowels

The vowel frequency charts below were devised from recordings made of three men from London, aged 55, 63 and 67 at the time of recording, reading the vowels in the context /h-d/. This is the environment chosen by Wells in his 1962 study (see II. Experimental procedure. Recording procedure) and it was adopted by Hawkins & Midgley (2005: 185). As Wells says, “The frame /h-d/ is particularly suitable for studies of English vowels, since (i) /h/ has so little influence on following vowels, and (ii) it so happens that a real English word results for nearly every ‘pure’ vowel in this sequence.” The words recorded were the following.

1. heed 2. hid 3. head 4. had
5. hard 6. hod 7. hoard 8. hood

To record instances of schwa in final position, where it may be particularly open, the word header was added. The participants were asked to repeat each word three times so that averages could be calculated from the three tokens for each vowel. This also compensates for beginning and end-of-list effects in reading (see Hawkins & Midgley 2005: 185).

The gaps in figure 3 mean that the reading taken was obviously inaccurate, either because the recording was too quiet or because of the presence of excessive creak.

Figure 4 was constructed from the averages recorded in figure 3 by using PLOTFORMANT.
Figure 3. The vowel formant frequencies of three male Londoners

<table>
<thead>
<tr>
<th></th>
<th>Steve Wood</th>
<th>Tony Corsini</th>
<th>Tony Saward</th>
<th>Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>age 55, Deptford (SE8)</td>
<td>age 67, Paddington (W2)</td>
<td>age 63, Barnes (SW13)</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>F2</td>
<td>F1</td>
<td>F2</td>
<td>F1</td>
</tr>
<tr>
<td>299</td>
<td>2507</td>
<td>288</td>
<td>2335</td>
<td>321 2297</td>
</tr>
<tr>
<td>331</td>
<td>2582</td>
<td>321</td>
<td>2312</td>
<td>268 2289</td>
</tr>
<tr>
<td>[i]</td>
<td>420 402 2285 2412</td>
<td>[i] 346 344 2126 2147</td>
<td>[i] 389 362 2142 2105</td>
<td>369 2221</td>
</tr>
<tr>
<td>388</td>
<td>2520</td>
<td>335</td>
<td>2169</td>
<td>362 2050</td>
</tr>
<tr>
<td>400</td>
<td>2433</td>
<td>353</td>
<td>2146</td>
<td>336 2123</td>
</tr>
<tr>
<td>562</td>
<td>2186</td>
<td>447</td>
<td>2026</td>
<td>476 1856</td>
</tr>
<tr>
<td>528</td>
<td>2132</td>
<td>470</td>
<td>1979</td>
<td>488 2025</td>
</tr>
<tr>
<td>692</td>
<td>1903</td>
<td>645</td>
<td>1757</td>
<td>673 1764</td>
</tr>
<tr>
<td>739</td>
<td>1859</td>
<td>681</td>
<td>1830</td>
<td>661 1851</td>
</tr>
<tr>
<td>[a:]</td>
<td>700 674 1127 1133</td>
<td>[a:] 609 613 1060 1071</td>
<td>[a:] 695 663 891 1021</td>
<td>650 1075</td>
</tr>
<tr>
<td>645</td>
<td>1140</td>
<td>606</td>
<td>1077</td>
<td>603 1195</td>
</tr>
<tr>
<td>678</td>
<td>1134</td>
<td>625</td>
<td>1078</td>
<td>691 979</td>
</tr>
<tr>
<td>[n]</td>
<td>601 584 879 918</td>
<td>[n] 586 574 968 989</td>
<td>[n] 649 896</td>
<td>602 934</td>
</tr>
<tr>
<td>570</td>
<td>935</td>
<td>528</td>
<td>1033</td>
<td></td>
</tr>
<tr>
<td>583</td>
<td>942</td>
<td>608</td>
<td>966</td>
<td></td>
</tr>
<tr>
<td>[o:]</td>
<td>414 412 683 664</td>
<td>[o:] 446 428 600 627</td>
<td>[o:] 475 472 660 660</td>
<td>437 650</td>
</tr>
<tr>
<td>387</td>
<td>659</td>
<td>405</td>
<td>600</td>
<td>448</td>
</tr>
<tr>
<td>436</td>
<td>650</td>
<td>433</td>
<td>682</td>
<td>493</td>
</tr>
<tr>
<td>431</td>
<td>1041</td>
<td>335</td>
<td>1066</td>
<td></td>
</tr>
<tr>
<td>385</td>
<td>1020</td>
<td>342</td>
<td>1004</td>
<td></td>
</tr>
<tr>
<td>411</td>
<td>1414</td>
<td>297</td>
<td>1408</td>
<td>456 1500</td>
</tr>
<tr>
<td>360</td>
<td>1267</td>
<td>300</td>
<td>1368</td>
<td>567 1555</td>
</tr>
<tr>
<td>[a]</td>
<td>662 652 1341 1324</td>
<td>[a] 714 717 1311 1305</td>
<td>[a] 758 758 1390 1490</td>
<td>709 1373</td>
</tr>
<tr>
<td>632</td>
<td>1334</td>
<td>719</td>
<td>1335</td>
<td>758 1554</td>
</tr>
<tr>
<td>663</td>
<td>1298</td>
<td>719</td>
<td>1270</td>
<td>758 1527</td>
</tr>
<tr>
<td>[ɔː]</td>
<td>537 528 1408 1353</td>
<td>[ɔː] 447 470 1481 1510</td>
<td>[ɔː] 537 499 1364 1493</td>
<td>499 1452</td>
</tr>
<tr>
<td>501</td>
<td>1369</td>
<td>451</td>
<td>1499</td>
<td>510 1581</td>
</tr>
<tr>
<td>546</td>
<td>1282</td>
<td>514</td>
<td>1550</td>
<td>452 1535</td>
</tr>
<tr>
<td>[ɔ]</td>
<td>583 599 1613 1593</td>
<td>[ɔ] 593 573 1512 1551</td>
<td>[ɔ] 593 604 1366 1585</td>
<td>592 1576</td>
</tr>
<tr>
<td>628</td>
<td>1545</td>
<td>563</td>
<td>1581</td>
<td>575 1645</td>
</tr>
<tr>
<td>588</td>
<td>1622</td>
<td>563</td>
<td>1562</td>
<td>646 1745</td>
</tr>
</tbody>
</table>
**Figure 4.** Cockney vowel formant frequencies. Stressed vowels and schwa

**Figure 5.** Cockney vowel formant frequency averages (stressed vowels and schwa) compared to formant frequencies for RP (relatively) pure vowels (in citation form) given in Cruttenden (2008: 99, Gimson 7th ed.), and Wells (1962), for male speakers in all cases

<table>
<thead>
<tr>
<th>Vowels of Cockney and RP</th>
<th>Averages for Cockney</th>
<th>Figures from Cruttenden (Gimson)</th>
<th>Figures from Wells 1962</th>
<th>Observations on Cockney Vs as compared to RP Vs</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>F1 311, F2 2389</td>
<td>F1 275, F2 2221</td>
<td>F1 285, F2 2373</td>
<td>lower, slightly fronter</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>F1 369, F2 2221</td>
<td>F1 382, F2 1958</td>
<td>F1 356, F2 2098</td>
<td>similar in height, fronter</td>
</tr>
<tr>
<td>/e/</td>
<td>F1 499, F2 2048</td>
<td>F1 560, F2 1797</td>
<td>F1 569, F2 1965</td>
<td>higher, fronter</td>
</tr>
<tr>
<td>/æ/</td>
<td>F1 679, F2 1825</td>
<td>F1 732, F2 1527</td>
<td>F1 748, F2 1746</td>
<td>higher, fronter</td>
</tr>
<tr>
<td>/ɑː/</td>
<td>F1 650, F2 1075</td>
<td>F1 687, F2 1077</td>
<td>F1 677, F2 1083</td>
<td>slightly higher, similar in frontness</td>
</tr>
<tr>
<td>/ɒ/</td>
<td>F1 602, F2 934</td>
<td>F1 593, F2 866</td>
<td>F1 599, F2 891</td>
<td>very slightly lower, fronter</td>
</tr>
<tr>
<td>/ɔː/</td>
<td>F1 437, F2 650</td>
<td>F1 453, F2 642</td>
<td>F1 449, F2 737</td>
<td>slightly higher, similar in frontness</td>
</tr>
<tr>
<td>/ɑ/</td>
<td>F1 391, F2 1073</td>
<td>F1 414, F2 1050</td>
<td>F1 376, F2 950</td>
<td>similar in height, fronter</td>
</tr>
<tr>
<td>/uː/</td>
<td>F1 387, F2 1438</td>
<td>F1 302, F2 1131</td>
<td>F1 309, F2 939</td>
<td>lower, fronter</td>
</tr>
<tr>
<td>/ʌ/</td>
<td>F1 709, F2 1373</td>
<td>F1 695, F2 1224</td>
<td>F1 722, F2 1236</td>
<td>similar in height, fronter</td>
</tr>
<tr>
<td>/ɜː/</td>
<td>F1 499, F2 1452</td>
<td>F1 513, F2 1377</td>
<td>F1 581, F2 1381</td>
<td>higher, fronter</td>
</tr>
<tr>
<td>/ə/</td>
<td>F1 592, F2 1576</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The figures given in Cruttenden (2008) are taken from Deterding (1997). No figures are given for /ɑ/, whose quality varies according to the phonetic environment, and whose average values may be taken to be equivalent to those for /ɔ/.

5. Conclusion

The last column of figure 5 compares the Cockney vowels produced by the three male speakers in this experiment with the results obtained for RP vowels in male speakers by Deterding and Wells. By way of conclusion, it would be useful to compare the outcome of the experiment described in this paper with previous observations made in the literature on the vowels of London speech.

Regarding the KIT vowel, it is generally assumed that it can be more central than in RP, but it was actually found to be fronter. The PALM vowel was not found to be fully back and low, as it may be in some accents, but slightly higher and similar in frontness to RP. The STRUT vowel was similar in height to RP and not lower, despite my anticipating a much lower articulation, as predicted in the literature, by using the symbol [a]. The LOT vowel was not found to be higher than in RP, as is often claimed, but very slightly lower and fronter.

References

SOME CONTROVERSIES ABOUT /v/ IN SERBIAN, TRANSFER IN ENGLISH, AND OTHER RELATED ISSUES

MAJA MARKOVIĆ AND BOJANA JAKOVLJEVIĆ

Outline

The sound /v/ has not aroused disputes in Serbian phonology and phonetics alone, but has also been known as a problematic sound across other languages. It has been called ‘hybrid’ (Hamann 2006) and even ‘schizophrenic’ (Kavitskaya 1999) in other languages, and not without a reason. The gist of the problem is the fact that it can be classified either as a labiodental fricative or as a labiodental approximant in a number of languages. It exhibits unexpected phonological behaviour and its acoustic character is rather unreliable and even arbitrary. Serbian is one of the languages where the status of /v/ is a linguistic battlefield, whereas in English, it is undisputedly a voiced fricative. In this paper, we try to tackle some of these problems in Serbian and English on the basis of phonological facts, the acoustic analysis of recorded material by L1 speakers of both languages and the presence of transfer in L2 (English) by native speakers of Serbian.

1. Introduction

The problem of the sound denoted by the IPA symbol /v/ has attracted the attention of phonologists and phoneticians apparently out of proportion to its size. To quote just some of the remarks found in the literature in relation to this segment: it has been compared to the ‘roar of a mouse’ (Padgett 2002), it has been called ‘hybrid’ (Hamann 2006), ‘schizophrenic’ (Kavitskaya 1999), accused of exhibiting ‘double-faced’ or ‘Janus-faced’ behaviour (Barkanyi & Kiss 2007) and complimented that it “has played a key role in discussions about abstractness in phonology […], and about the larger organization of phonology” (Padgett 2002). Although it is regarded as just another fricative in English, the phonologists and phoneticians
Some Controversies about /v/ in Serbian, Transfer in English

dealing with Serbian still have not made up their minds as to how to classify this segment. The difference in the articulation of /v/ in the two languages, despite its common IPA symbol, has also been an everyday problem for the learners of English whose L1 is Serbian. In this paper, we shall point to some of the problems regarding the nature of /v/ cross-linguistically, give the results of our findings based on the acoustic analysis of English and Serbian recorded material by respective native speakers, and propose the analysis of this segment within a wider phonological perspective.

2. Problems

Phonological systems of a number of languages have a two-way contrast between /f/ and /v/, and most linguistic systems will recognize this contrast as a pair of a voiceless and voiced labiodental fricative. The reasons for doing so are both phonological and phonetic. In terms of phonological distinctions, the members of this pair regularly enter predictable phonological processes like all other voiceless/voiced fricative pairs. These processes include:

- final obstruent devoicing, where the voiceless member /f/ only occurs in the word final position in German, Russian (Hamann & Sennema 2005), some dialects of Slovene (Jurančič Petek 2009), and in other languages;
- processes of regressive or progressive voicing assimilation;
- phonotactic rules of typical obstruent/sonorant occurrence.

Among the phonetic arguments for using this classification, the fricative realization implies the presence of turbulences in the soundwave of the fricative.

Yet, a number of languages fail to comply with the above listed rules. Within the phonological systems of some languages, word final devoicing can take place, and the sound may participate in regressive voicing assimilation, but fail to do so in progressive assimilation. It can also undergo voicing assimilation, but fail to trigger it (e.g. Russian). The latter behaviour speaks in favour of analyzing /v/ as a sonorant rather than a fricative. Moreover, in some languages, /v/ can occupy positions in the syllable which are typically reserved for sonorants. In other languages (e.g. Norwegian), it can occupy the positions exclusively intended for sonorants and the positions exclusively intended for obstruents.

To top up the arguments for the weird behaviour, the acoustic analysis of /v/ also reveals a rather messy situation, where it can have a typically fricative nature, a typically sonorant nature or even resemble a plosive.
The list of problems can be further extended by the possibility of word final vocalization of /v/, where it changes into the labio-velar semivowel /w/ and further into a back vowel proper (e.g. Slovak and some dialects of Slovene).

The problems listed here are found across various languages, and typically involve some kind of relation between /f/ : /v/ : /w/ and ø. Processes involving some kind of vacillation between these segments are attested in a number of languages, both synchronically and diachronically. If a process involves the change in the rightward direction, we can speak of ‘fortition’; if it moves leftwards, the process in question is referred to as ‘lenition’.

If we take as an example the final occurrence of /v/ cross-linguistically, we can see that it is up to a particular language whether to choose (a) fortition i.e. the occurrence of the voiceless fricative /f/- as in German or Russian; (b) lenition – sonorization which can ultimately lead to the loss of a segment – as in Slovak and some dialects of Slovene; or (c) to do nothing, and leave it as a (disputably) voiced fricative – as is the case of English.

One thing worth mentioning is that very few languages have a three-way contrast between the above segments. In our opinion, this simple truth may be crucial for positing some of the solutions to the problem of /v/. It is, however, rarely brought up in the literature. Hamann & Sennema (2005) make this valuable remark: “A three-way distinction of labiodentals is crosslinguistically very unusual. Apart from Dutch, we know only of two other languages that have the same three labiodental categories, namely the Edoid languages Isoko and Urhobo, spoken in Nigeria”, quoting Ladefoged & Maddieson (1996).

The strange phonological behaviour led linguists to try to find explanation for this ambiguous behaviour in the articulatory/acoustic character of /v/. This is where the hell broke loose, because it turned out that a number of languages actually did not have a fricative realization of /v/ as it had long been presumed.

Of course, the story needed an explanation. Some of the propositions included the account that /v/ was ‘underlyingly’ /w/, although in terms of its phonetic realization it might not be so. Padgett (2002) proposed an entirely new feature, called ‘narrow approximant’, which is potentially distinctive for the languages of the world, no matter how rarely it may be used. On the whole, this seems a complex story without a solution.
3. Facts about /v/ in English and Serbian

3.1. English

English is one of the languages in which no one has challenged the fricative nature of /v/. In terms of its phonological behaviour, it undergoes voicing assimilation just like any other voiced fricative. For example, /v/ of the words ‘of’, ‘we’ve’ will change into /f/ in ‘of course’ or ‘we’ve found it’ (examples taken from Cruttenden 1994: 257).

In terms of its phonetic voicing, it is, like all other English obstruents, devoiced in the word final position and partially voiced word initially. According to Cruttenden (1994: 163), /v, ð, z, ð/ are fully voiced between voiced sounds; “in initial and (especially) in final positions, the voiced fricatives may be partially or almost completely devoiced; e.g. initially in van, that, zoo (…) only the latter part of the friction is likely to be voiced, and finally (…) the friction is typically voiceless, though the consonant remains lenis”.

As regards its position in the syllable, it also consistently only occurs in the positions typical of obstruents. It never occurs in consonant clusters, either initial or final (with the exception of /v/ being followed by a ‘post-final’ consonant, as in loved /λʌvd/ or loves /λʌvz/, but this is a legitimate position for an obstruent).

3.2. Serbian

In Serbian, /v/ is traditionally classified as a sonorant. The reasons are primarily phonological. For one thing, /v/ is known not to undergo voicing assimilation as voiced obstruents do, e.g.

(a) iz + tupiti > istupiti, od + kazati > otkazati
(b) lov + ca > lovca, ovas + en > ovse

Historically, /v/ was derived from a previous bilabial semivowel /w/. The voiceless fricative /f/ entered the sound system of Serbian much later, mainly through Turkish loanwords, and later via loanwords from English and other languages. This historic development partly explains lack of assimilation in words such as ovca.
Regarding its distribution, it is found in typically sonorant positions in initial consonant clusters, following obstruents in two or three consonant clusters: tvoj, dva, kvar, gvožđe, stvar. Apart from /m/, /v/ is the only sonorant which can be the first element of initial two-consonant clusters: vreme, vлага, etc. Word final consonant sequences are generally felt as foreign in Serbian, but it should be noticed that some sequences are quite readily accepted, such as in the words takt, keks, princ, film, disk, saft, kamp, šund, dizajn; whereas sequences involving /v/ are very infrequent (as in gotovs, nerv, but it should be noted that the distribution in both of words speaks in favour of an obstruent analysis, being equivalent to indeks, gips, kolaps; and park, bard, punč, respectively).

Another reason why /v/ is commonly classified as a sonorant is based on the articulatory and acoustic properties of this sound. Although it may undergo devoicing in certain positions, it generally has the characteristics of approximant articulation. In terms of its articulation, the contact between the upper teeth and the lower lip is rather loose, and the energy seems too low to produce friction. From the acoustic standpoint, it is characterized by low energy output, barely visible friction (if any), formant structure, and, usually, by the presence of voicing.

The majority of textbooks on Serbian and Croatian phonetics and phonology traditionally analyze /v/ as a sonorant. In some of the more recent works the sonorant status of /v/ has been challenged, and the authors such as Gudurić & Petrović (2006), as well as Subotić (2005) speak in favour of a fricative, particularly referring to the opposition /f/ : /v/, which functions as a distinctive opposition in contemporary Serbian. The acoustic investigations reported in Gudurić & Petrović (2006) are indicative of a highly idiosyncratic (and disputably fricative) articulation

---

1 stv- seems to be the only common three consonant cluster with /v/ as its third element. Words such as ‘zdvojiti’, listed in the dictionary Rečnik Matice srpske, are not commonly heard in the language. The sequence skv- is only found in non-standard or foreign words, such as ‘skvičati’, ‘skvo’, and a few more. /v/ generally has a much more restricted occurrence as the third element of a three-consonant cluster, compared to /r/, the other post-initial sonorant found in this position.

2 For a detailed account of /v/ in Serbian and Croatian by various phoneticians and phonologists, see Gudurić & Petrović (2006). Among those not mentioned in this paper, we would mention Bakran (1996), who gives phonetic evidence for sonorant realization in Croatian, and Jelaska (2004), who analyzes /v/ as a sonorant from the phonological point of view.
of /v/ in various contexts, but the authors nonetheless take the position that the default realization of /v/ is fricative.

4. Acoustic analysis

The experimental part of the paper involved the acoustic analysis of English and Serbian /v/ as well as the analysis of English /v/ produced by Serbian native speakers. The subjects were two native speakers of English and two native speakers of Serbian (all males) who had a minimum of twelve years of learning the English language. The corpus consisted of English and Serbian words/phrases illustrating /v/ in a variety of phonological contexts: word-initially, intervocally, word-finally, preceding both voiced and voiceless consonants as well as in a post-consonantal position. The subjects were recorded in the soundproof room at the Faculty of Philosophy in Novi Sad with the sampling rate of 44,100 Hz. English subjects were asked to pronounce the English tokens only, whereas Serbian subjects were recorded pronouncing both Serbian and English tokens with a short pause made between the two sets of recording. The acoustic analysis was done in Praat (version 5.0.42) with the focus on the following features: (a) the presence and distribution of aperiodic energy/periodic energy with the significant increase in amplitude, (b) the presence and average values of the formant frequencies, and (c) the presence and duration of voiced phonation.

4.1. English /v/

The corpus contained the following English words: visa, velar, viva, Venus, vegan, vodka, volley, volume, vomit, vox, lava, bravo, seven, seventh, prevail, even, proverb, novel, hovering, moving, grieve, leave, Steve, reeve, sleeve, of, groove, prove, move, remove, love bite, dive-bomb, love potion, have problems, proved, gravedigger, love-token, leave-taking, have got, brave girl, live concert, of course, have vanished, have vended, waveform, brave face, save that, prove this, brave thing, drive-thru, wives, lives, love-song, gravestone, love genre, brave gigolo, slave ships, live show, love-hate, leave home, have judged, have joked, love child, love children, caveman, movement, love nest, have-not, loveless, lovely, graverobber, have risking, brave woman, driveway, prove useful, graveyard.

The acoustic analysis shows that the English phoneme /v/ has the undisputable status of a non-sibilant fricative in all phonological environments. It is characterized by the obligatory presence of low-
intensity friction extending up to 22,000 Hz and its spectra are diffuse with energy relatively evenly distributed over a wide range of frequencies. The aperiodic energy of slightly higher intensity is observed in the range of approximately 5,000 to 14,000 Hz, although its concentration tends to be 200-500 Hz lower on the frequency scale in front of a rounded vowel. Due to the fact that the presence of lip-rounding lengthens the front resonator, the friction tends to be generated at slightly lower frequencies in such contexts.

Similarly to other voiced obstruents, English fricative /v/ most often exhibits full phonetic voicing in intervocalic position (Figure 1-a). Short interruptions of the pitch contour were observed in only 5 out of 20 recorded tokens, most often in the middle of the constriction interval. In addition, the formant structure tends to be visible, which is especially the case if the following vowel is stressed.

**Figure 1.** Spectrograms and waveforms of the English words (a) *prevail* and (b) *vegan* with the segment /v/ highlighted

However, interestingly enough, English /v/ tends to be produced with significant vocal fold activity in word-initial position as well (Figure 1-b). The measurements based on the recordings of two English subjects show that the voice bar coincides with almost entire constriction interval, thus pointing to almost full phonetic voicing (Table 1). Since this behavior was not observed in the production of other voiced obstruents in word-initial position, it can only be attributed to the peculiar nature of /v/ and its
inclusion towards phonation and weakening. Similarly to its intervocalic realizations, English word-initial /v/ is also characterized by visible formant structure. The behavior of F1 tends to be rather stable regardless of the vocalic environment, whereas the average F2 frequency tends to be significantly lower in front of a back vowel (Table 2). The average frequency of F3 was not measured because it was rather difficult to track it in most of the analyzed tokens.

Table 1. The average duration of the voiced interval in relation to the constriction interval (ms) for English word-initial /v/ produced by two recorded subjects

<table>
<thead>
<tr>
<th></th>
<th>Subject 1</th>
<th>Subject 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constriction</td>
<td>114.9</td>
<td>101.2</td>
</tr>
<tr>
<td>Voice bar</td>
<td>102.9</td>
<td>85</td>
</tr>
<tr>
<td>Average</td>
<td>89.56%</td>
<td>83.99%</td>
</tr>
</tbody>
</table>

Table 2. The average values of F1 and F2 (Hz) for English word-initial /v/ in front of /i:/ and /d/.

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>256</td>
<td>1707</td>
</tr>
<tr>
<td>/d/</td>
<td>268</td>
<td>960</td>
</tr>
</tbody>
</table>

The vocal fold activity characterizing word-final /v/ tends to be less uniform and rather subject-dependent. While one subject produces it as partially voiced, the production of the other subject shows significant reduction of the vocal fold activity, occasionally exhibiting even complete devoicing (Table 3).

Table 3. The average duration of the voiced interval in relation to the constriction interval (ms) for English word-final /v/ produced by two recorded subjects

<table>
<thead>
<tr>
<th></th>
<th>Subject 1</th>
<th>Subject 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constriction</td>
<td>161.1</td>
<td>175.3</td>
</tr>
<tr>
<td>Voice bar</td>
<td>80.8</td>
<td>28.1</td>
</tr>
<tr>
<td>Average</td>
<td>50.15%</td>
<td>16.03%</td>
</tr>
</tbody>
</table>

In the latter case, the amount and intensity of friction tend to be slightly greater in comparison to word-initial realizations, whereas formants tend to be lost in both cases. Higher intensity of friction probably results from the lack of the vocal fold activity, which causes the loss of balance in the air
pressure below and above the glottis. This significant output of aperiodic energy consequently leads to the lack of formant structure.

The voicing properties of pre-consonantal /v/ depend on the phonetic voicing characterizing the consonant in question. Pre-consonantal /v/ followed by a voiceless obstruent is commonly subjected to regressive assimilation and as such, it is characterized by the absence of formant structure. Complete devoicing is most often observed in front of /θ/ and /ʃ/ (Figure 2-a). However, if the articulation of the following obstruent is at a considerable distance from the labiodental region such as the one of /s/, /ʃ/ or /tʃ/, the process of devoicing is not significant (Figure 2-b). As for other voiceless obstruents, the presence of regressive assimilation is strongly dependent on the carefulness of pronunciation as well as the frequency of use of the analyzed token. Accordingly, the process of devoicing of pre-consonantal /v/ is much more common in fixed phrases or compounds such as e.g. of course or love-hate than in less frequently used ones such as e.g. live concert or leave home, which tend to be pronounced more carefully.

**Figure 2.** Spectrograms and waveforms of the English words (a) drive-thru and (b) gravestone with the segment /v/ highlighted

If the following consonant is voiced, English pre-consonantal /v/ most often exhibits full phonetic voicing (Figure 3-a). The interruptions of the pitch contour as well as the occasional process of complete devoicing are sometimes observed in front of voiced consonants whose articulation involves the formation of complete closure such as stops and nasals.
(Figure 3-b) as well as in front of the fricative /ð/. However, the formant patterns are most often not clearly identifiable in either case.

**Figure 3.** Spectrograms and waveforms of the English strings (a) *love genre* and (b) *caveman* with the segment /v/ highlighted

### 4.2. Serbian /v/

The corpus of Serbian words included: *viza, vila, vivak, vino, višak, votka, voli* (from *voleti*, 3rd person sg.), *volim* (from *voleti*, 1st person sg.), *vodnik, volt, lava, bravo, neven, sevap, privid, Ivan, provod, novac, uvući, muva, kriv, liv, Stiv, živ, siv, nov, suv, gluv, gruv, nagluv, suv barut, sav bitan, protivpožarni, zov prirode, ovde, pravda, jevtin, protivteža, kavga, kavgadjija, olovak, nabavka, protivvazdušni, suv vazduh, siv fenjer, dirljiv film, protivzakonit, živ zakopan, ovseni, pivski, nov život, nov žeton, bivša, bivši, prav hodnik, kriv hitac, dživdžan, dživdžani, ovčar, kovčeg, stidljiv đak, hrapav đon, Lovćen, nov čup, ovca, lovci, Brankov most, nov mantil, klovnovi, ovnovi, daždevnjak, ljubav njena, pavlaka, Pavle, divljak, divljina, uvreda, gavran, plav jastuk, surov jezik, hvala, tvoj, dvorac, tvorac, kvar.

The behavior of the Serbian phoneme /v/ appears to be much less uniform, both in terms of voicing and the manner of articulation. The spectral image of Serbian /v/ tends to vary from one speaker to another and this is the case in all positions within a word, except when used intervocally.

Considering word-initial realizations, one subject produces it as a clear sonorant. The friction is completely absent, the sound wave is periodic and
it often assumes the pattern of the following vowel, which makes it rather difficult to establish a clear boundary between them (Figure 4-a). As it is expected, the formants are clearly visible, with their average frequencies adhering to the same pattern as in English. Accordingly, F1 tends to be rather stable regardless of the vocalic environment, whereas F2 tends to be lower on the frequency scale if the following vowel is back (Table 4). However, the second subject produces word-initial /v/ as a rather weak non-sibilant fricative (Figure 4-b). The friction characterized by relatively even distribution over a wide range of frequencies is most often observed in the second half of the constriction interval, whereas its intensity tends to be significantly lesser in comparison to the corresponding word-initial realizations in English. The only property they have in common is related to the presence of phonation. Thus, both subjects produce word-initial /v/ as fully voiced in all the analyzed tokens.

**Figure 4.** Spectrograms and waveforms showing the first syllable of the Serbian word *vino*, produced by (a) the first and (b) the second Serbian subjects.

**Table 4.** The average values of F1 and F2 (Hz) for Serbian word-initial /v/ in front of /i/ and /o/.

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th></th>
<th>F2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>/i/</td>
<td>/o/</td>
<td>/i/</td>
<td>/o/</td>
</tr>
<tr>
<td>308</td>
<td>331</td>
<td>1689</td>
<td>922</td>
<td></td>
</tr>
</tbody>
</table>
The behavior of word-final /v/ tends to be much more uniform in relation to the manner of articulation. Both subjects produce it as a non-sibilant fricative and, as in English, the formant structure tends to be lost. What they differ in is the presence of phonetic voicing. While one subject significantly reduces the vocal fold activity with the occasional presence of complete devoicing, the production of another subject exhibits full phonetic voicing in most of the analyzed tokens (Table 5). However, what appears to be interesting is the process of lenition occasionally taking place in the latter case. The friction of extremely low intensity is present only in the first half of the constriction interval and then it dies out, which is the point when the pattern of the sound wave becomes regular, the amplitude starts to show gradual increase and the formants start to be visible (Figure 5). This is especially the case when the preceding vowel is /u/.

**Figure 5.** Spectrogram and waveform of word-final /v/, extracted from the Serbian word *suv*

| Table 5. The average duration of the voiced interval in relation to the constriction interval (ms) for Serbian word-final /v/ produced by two recorded subjects |
|---|---|
| **Constriction** | **Voice bar** | **Constriction** | **Voice bar** |
| Subject 1 | Subject 2 |
| 154.4 | 28 | 152.33 | 142 |
| 18.13% | | 93.21% | |
If followed by a voiceless consonant, Serbian /v/ shows the same kind of erratic behavior as in the case of word-initial realizations. It behaves either as a sonorant or a non-sibilant fricative depending on the subject. However, what can be observed in both cases is the lack of tendency towards regressive assimilation (Figure 6-a). Both subjects tend to produce it as voiced throughout the entire constriction interval in most of the analyzed tokens. What is equally important, the lack of regressive assimilation causes Serbian pre-consonantal /v/ to be almost identical in front voiceless and voiced consonants (cf. Figure 6-a and 6-b). The only difference is a slight reduction in the amount of friction and possibly the presence of rather faint formant structure, occasionally observed in the latter case.

**Figure 6.** Spectrograms and waveforms of the Serbian words (a) kovčeg and (b) dživdžan with the segment /v/ highlighted

The only position in which Serbian phoneme /v/ tends to exhibit rather uniform behavior is the intervocalic position. Intervocalic /v/ is voiced throughout the entire constriction interval and, even more importantly, it is most often produced as a clear sonorant. The production of both subjects shows the complete absence of friction, the regular pattern of the sound wave and clearly visible formant structure (Figure 7). For this very reason, it is rather difficult to establish a clear-cut boundary between the segment and the surrounding vowels.
Summarizing these observations, the behavior of Serbian /v/ tends to vary both in relation to a speaker and phonological context. Considering speaker-dependent variations, Serbian /v/ can be realized either as a sonorant or a non-sibilant fricative characterized by significant phonation which is most often not reduced either word-finally or in front of a voiceless consonant. However, it is interesting to notice that in the former case, Serbian /v/ behaves as a fricative if used word-finally, whereas in the latter case, it shows a strong tendency towards sonority and weakening if used intervocalically. The same behavior is observed in the case of post-
consonantal realizations where Serbian /v/ behaves as a clear sonorant (Figure 8).

### 4.3. Transfer into English

The presence of transfer from Serbian into English was examined only on those properties that both Serbian subjects share and which were not observed in English. They include: (a) the regressive assimilation in front of voiceless obstruents and (b) the manner of articulation in intervocalic position.

**Figure 9.** Spectrogram and waveform of the English word *waveform*, produced by the Serbian subject. The sequence /v# f/ is highlighted.

Serbian subjects tend to perceive the difference between Serbian and English /v/ in front of voiceless obstruents and accordingly they reduce the vocal fold activity, which results in the absence of the expected transfer. As it can be observed in Figure 9, the syllable-final /v/ of the English word *waveform* tends to be produced without any phonation, which makes it rather difficult to establish a boundary between /v/ and the following /f/. However, the transfer can be observed in the production of English intervocalic /v/. As in Serbian, it is most often produced as a clear sonorant; the friction is completely absent, the sound wave is periodic and the formants are clearly visible (Figure 10).
5. Conclusion

In Serbian, there are generally no strong phonological arguments to support the claim that /v/ is a fricative, as it apparently does not enter voicing assimilation, and Serbian lacks word-final voicing neutralization. In phonetic terms, the results of the research point to a rather idiosyncratic articulation of the Serbian phoneme /v/. If we assume that its default realization is sonorant, the process of fortition takes place word-finally, resulting in rather weak fricative realizations. On the other hand, if the default realization is the one of a fricative, the lenition is systematically observed intervocally as well as in post-consonantal position, which results in sonorant realizations.

According to Johnson (2003) and Ohala (1983), such arbitrary behavior can be accounted for by aerodynamic ‘incompatibility’ of voicing and friction/noise which results in the surface instability of [v]. This is especially true of poorly cued and aerodynamically unfavorable positions (i.e. in the vicinity of consonants and/or word-boundary), but also in lenition triggering intervocalic position. The periodic and aperiodic energy released in the articulation of a given consonant appear to be on a

---

3 It should be noted that this only refers to Standard Serbian, while voicing assimilation of /v/ is attested in a number of other dialects.
seesaw and their surface realizations in particular phonological contexts seem to be in the domain of the physical and physiological.

Another important factor worth mentioning in the system of the Serbian language is perception. Standing in a two-way opposition with /f/, /v/ is in no danger of being confused with any other phoneme, even if underarticulated, which is not the case of the languages with a three-way distinction, be it either /f/ : /v/ : /u/ (as in Dutch), or /f/ : /v/ : /w/ (as in English).

Such a difference in the phonetic realization of /v/ and the different phonological function it has in the systems of Serbian and English generally result in the transfer observed in the learners of English whose native language is Serbian. The transfer is commonly found in intervocalic position, where the consonant tends to be articulated as a clear sonorant. In other positions, it shows a strong tendency towards voicing, resulting in rather weak friction, especially in comparison to its English equivalent. In relation to this matter, the native speakers of English often report perceiving /v/ in the production of Serbian speakers as ‘loose’, ‘short’, ‘not strong enough’, which also indicates the presence of transfer in the learners’ interlanguage. Accordingly, in the process of acquiring English phonology for L2 learners of English, it is important to realize a three-way contrast between /f/ : /v/ : /w/, and keep the three well apart by producing the voiced fricative with more turbulence than the Serbian sound.

References

Some Controversies about /v/ in Serbian, Transfer in English

VOT TRANSFER IN THE PRODUCTION OF ENGLISH STOPS BY SERBIAN SPEAKERS

BOJANA JAKOVLJEVIĆ

Outline

The aim of the paper is twofold: firstly, to measure the duration of the Voice Onset Time (VOT) characterizing British English and Serbian word-initial stops at the beginning of a stressed syllable; and secondly, to examine the presence and degree of transfer in the production of English stops by Serbian native speakers. The subjects are two native speakers of English and five native speakers of Serbian (all females) who have learned English as their L2 during the last 12 years. The corpus consisted of Serbian-English pairs of words illustrating each stop before Serbian /i/ (long-rising accent) vs. English /i:/ and Serbian /o/ (short-falling accent) vs. English /ɒ/ respectively. Despite the fact that the Serbian subjects have received considerable training in English pronunciation, the results of the research point to the strong presence of transfer in the production of English word-initial stop consonants. This is particularly the case in the articulation of voiced stops, whereas the production of the corresponding voiceless equivalents tends to be under lesser influence of the subjects’ L1, occasionally exhibiting the presence of phonetic hypercorrection.

1. Introduction

The paper presents the results of the measurements of the Voice Onset Time (VOT) characterizing British English and Serbian word-initial stops at the beginning of a stressed syllable and examines the presence and degree of transfer in the production of English stops by Serbian native speakers. Accordingly, the introductory part of the paper has two most significant goals: firstly, to define the mechanism of transfer; and secondly, to provide a brief account of the concept of the VOT.

The tendency of learners to apply the knowledge of L1 in the process of mastering the system of L2 is known under the name of transfer.
According to the Contrastive Analysis Hypothesis (CAH) where the concept was originally proposed as so-called interference, in the process of learning a foreign language, the system of L2 tends to be filtered through the system of L1 with L1 facilitating the learning of L2 when the target structures are similar and interfering with it when they are different or completely absent (Celce-Murcia, Brinton and Goodwin 1996: 19-20). The two situations are nowadays commonly referred to as a positive and negative transfer respectively.

The CAH as well as its approach towards the notions of interference or transfer have been challenged for a number of reasons. The most significant ones are their inability to accurately predict the degree of difficulty learners will have with a particular structure (i.e. subtle differences between the structures of L1 and L2 are sometimes more difficult to acquire than the more prominent ones) as well as the inability to predict the directionality of difficulty (i.e. whether it will be more difficult for the native speakers of one language to master a dissimilar or non-existent structure in another, or vice versa). However, despite the aforementioned criticism, the concept of transfer is accepted as valid by most researchers in the field. Equally importantly, its mechanisms can be observed at all levels of linguistic analysis, including phonetics and phonology.

Shifting the focus to the production of stop consonants, their voicing properties are acoustically defined in terms of the Voice Onset Time (VOT), which is the interval beginning with the stop release and ending with the start of modal voicing of the following vowel (Ladefoged 2001: 120). Considering the scale of possible VOTs, stops are commonly classified into three major categories: voiced, voiceless unaspirated and voiceless aspirated. If the vocal fold vibration starts before the release stage, a given stop is characterized by negative VOT and it is said to be voiced. In the production of voiceless unaspirated stops, the voiced phonation of the following vowel starts more or less immediately after the release, which results in zero VOT. Finally, positive VOT is the indication that a given stop is aspirated because the presence of aspiration has the potential to delay the voicing of the following vowel.

What is equally important, the VOT is a means of making the opposed members of the [+/- voiced] contrast as distinct as possible so that they can be perceptually differentiated without difficulties. As Clark & Yallop (1990: 90) point out, in the languages whose phonologically voiced stops do not exhibit significant phonetic voicing such as English and German, the corresponding voiceless counterparts will tend to be produced with long positive VOT. However, if the phonologically voiced stops of a given
language do not show significant discrepancies between the concepts of phonological and phonetic voicing such as French and Dutch, the corresponding voiceless equivalents will not be considerably aspirated (Clark & Yallop 1990: 90). This general tendency of languages to rely on the VOT in order to make the members of the phonological opposition as different as possible was originally suggested by Keating (1984: 286-319) as the Polarization principle.

Considering the fact that the world languages use different VOT values in order to mark the voicing contrast as well as the fact that the mechanism of transfer works at all linguistic levels including the phonological one, it is expected that the learners of a given foreign language will produce its stops with the VOT values which characterize the corresponding stops in their native language. In other words, they are expected to rely on those phonetic cues which are responsible for distinguishing between phonologically voiced and voiceless stops in their L1.

2. The goals and methodology of the research

As already mentioned, the goal of the paper is twofold: firstly, to measure the VOT characterizing British English and Serbian word-initial stops in front of a stressed vowel; and secondly, to explore the presence and degree of transfer in the production of English stops by Serbian native speakers.

The subjects were two native speakers of English and five native speakers of Serbian (all females) whose speech contained no significant dialectal features. All Serbian subjects were at the time the graduate students at the Department of English Language and Literature at Novi Sad University. They have learned English as their L2 during the last 12 years and thus received considerable training in English pronunciation.

The corpus consisted of Serbian-English pairs of words illustrating each stop in front Serbian /i/ (long-rising accent)/English /i:/ and Serbian /o/ (short-falling accent)/English /o/ respectively (Table 1). Due to the fact that the duration of the VOT tends to vary with respect to the vocalic environment, the intra-linguistic choice of vowels in either language was determined by the differences in both the height and the part of the tongue involved in their articulation. The inter-linguistic selection resulted from both articulatory and acoustic similarities i.e. the average F1 and F2 frequencies of the chosen vowels in the languages in question.

In order to examine the presence of transfer, the Serbian subjects were asked to pronounce the lists of both Serbian and English words with a short break made between the two sets of recording, whereas the English
subjects were asked to pronounce the English words only. The recorded tokens were subjected to the acoustic analysis in Praat (version 5.0.42) with the average VOT values expressed in milliseconds (ms) and the relevant results expressed in percentage terms. Considering a relatively small number of subjects, it is important to point out that the obtained VOT values should not be considered as final. Accordingly, the research on a larger sample would probably yield slightly different values. However, the ratios between the obtained results as well as some general tendencies observed in the speech of either Serbian or English subjects should be considered as relevant due to their systematic occurrence.

Table 1. The list of Serbian-English pairs of words which were recorded and subjected to the acoustic analysis

<table>
<thead>
<tr>
<th>Stop</th>
<th>Following vowel</th>
<th>Serbian-English pairs of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>S /i/ (l-r)/E /i:/</td>
<td>piling-peeling</td>
</tr>
<tr>
<td></td>
<td>S /o/ (s-f)/E /o/</td>
<td>pot-pot</td>
</tr>
<tr>
<td>/t/</td>
<td>S /i/ (l-r)/E /i:/</td>
<td>Tina-teeny</td>
</tr>
<tr>
<td></td>
<td>S /o/ (s-f)/E /o/</td>
<td>top-top</td>
</tr>
<tr>
<td>/k/</td>
<td>S /i/ (l-r)/E /i:/</td>
<td>Kina-keener</td>
</tr>
<tr>
<td></td>
<td>S /o/ (s-f)/E /o/</td>
<td>konzul-consul</td>
</tr>
<tr>
<td>/b/</td>
<td>S /i/ (l-r)/E /i:/</td>
<td>bina-beanie</td>
</tr>
<tr>
<td></td>
<td>S /o/ (s-f)/E /o/</td>
<td>boks-box</td>
</tr>
<tr>
<td>/d/</td>
<td>S /i/ (l-r)/E /i:/</td>
<td>diva-diva</td>
</tr>
<tr>
<td></td>
<td>S /o/ (s-f)/E /o/</td>
<td>dok-dock</td>
</tr>
<tr>
<td>/g/</td>
<td>S /i/ (l-r)/E /i:/</td>
<td>gica-geese</td>
</tr>
<tr>
<td></td>
<td>S /o/ (s-f)/E /o/</td>
<td>golf-golf</td>
</tr>
</tbody>
</table>

3. The results of the research: VOT in Serbian and English

The phonemic inventories of Serbian and English share the same set of stop consonants. They occur in [+/- voiced] phonological opposition and include /b d g/ and /p t k/ respectively. However, the results of the measurements show that the two languages make use of rather distinct VOT values in order to mark the voicing contrast.

The production of Serbian word-initial /b d g/ involves considerable vocal fold activity coinciding with the closure, which results in rather long
negative VOT (Figure 1). According to the performed measurements, the average duration of the voiced interval beginning with the first glottal pulse and ending with the release of pressure is 115.23 ms. Despite the fact that the average values show the increase from /b/ and /d/ to /g/ (Table 2), the same continuum does not apply to each individual subject and thus should not be considered as universal. What tends to be universal is the influence of the vocalic environment on the voiced phonation. Namely, each phonologically voiced stop tends to have a slightly longer voiced interval in front of /i/ than in front of /o/ (Table 2) and such a tendency has been observed in the production of all the recorded subjects.

**Figure 1.** The spectrogram of the initial CV sequence of the Serbian word dok, produced by the female subject

![Spectrogram](image)

**Table 2.** The average duration of the voiced interval (ms) for Serbian word-initial /b d g/ followed by /i/ and /o/ at the beginning of a stressed syllable

<table>
<thead>
<tr>
<th></th>
<th>/i/</th>
<th>/o/</th>
<th>/i/</th>
<th>/o/</th>
<th>/i/</th>
<th>/o/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/b/</td>
<td>120</td>
<td>92.6</td>
<td>126.9</td>
<td>105.2</td>
<td>133.6</td>
<td>113.1</td>
</tr>
<tr>
<td>/d/</td>
<td>106.3</td>
<td>116.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/g/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>123.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>115.23</td>
</tr>
</tbody>
</table>
Despite the fact that they are often described as partially voiced (Yavas 2005: 57) or produced with some voicing lost when used word-initially (Collins & Mees 2003: 51), the results of the research point to complete devoicing of English /b d g/ in this position (Figure 2). The spectrograms of 24 recorded tokens (12 pronounced by each subject) show no presence of vocal fold vibration regardless of the place of articulation or the vocalic environment. Thus, the first acoustic manifestation of English word-initial /b d g/ is the release of pressure, whereas the phonation of the following vowel starts immediately after it, resulting in zero VOT.

**Figure 2.** The spectrogram of the initial CV sequence of the English word *dock*, produced by the female subject

Due to the fact that Serbian word-initial /b d g/ do not show significant discrepancies between the concepts of phonetic and phonological voicing, the corresponding voiceless counterparts are hardly audibly aspirated. The measurements point to relatively short positive VOT with the average duration of 32.05 ms. Similarly to the duration of the voiced intervals of Serbian word-initial voiced stops, the duration of the VOT increases from
the bilabial and (alveolar) dental\(^1\) to the velar and it tends to be slightly longer in front of /i/ than in front of /o/ (Table 3). However, differently from the behavior of the voiced stops, the relation between the VOT and the place of articulation does apply to each subject and, accordingly, it can be considered as universal. As Ladefoged (2003: 98) points out, the tendency of stops whose closure is formed further back in the oral cavity to have longer VOT is often observed cross-linguistically. The same goes for aspiration whose presence tends to be slightly stronger in front of /i/ (Ladefoged 2003: 5).

Table 3. The average duration of the VOT (ms) for Serbian word-initial /p t k/ followed by /i/ and /o/ at the beginning of a stressed syllable

<table>
<thead>
<tr>
<th></th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>23.3</td>
<td>33.3</td>
<td>48.6</td>
</tr>
<tr>
<td>/o/</td>
<td>19.2</td>
<td>27</td>
<td>40.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>/i/</th>
<th>/o/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>21.25</td>
<td>30.15</td>
</tr>
<tr>
<td>/o/</td>
<td></td>
<td>44.75</td>
</tr>
</tbody>
</table>

Table 4. The average duration of the VOT (ms) for English word-initial /p t k/ followed by /i:/ and /\d/ at the beginning of a stressed syllable

<table>
<thead>
<tr>
<th></th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>42.25</td>
<td>59.5</td>
<td>74.5</td>
</tr>
<tr>
<td>/\d/</td>
<td>34.25</td>
<td>49.75</td>
<td>61.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>/i:/</th>
<th>/\d/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>38.25</td>
<td>54.62</td>
</tr>
<tr>
<td>/\d/</td>
<td></td>
<td>67.87</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>53.58</td>
</tr>
</tbody>
</table>

What remains to be examined is the average VOT value characterizing English word-initial /p t k/. Considering the fact that the word-initial realizations of English phonologically voiced stops tend to be produced without any vocal fold activity, it is expected that the corresponding voiceless counterparts will be articulated with considerable aspiration, which is exactly the case. The measurements show that the average

\(^1\) In the articulation of Serbian /t d/, the tip of the tongue rises to touch partly the inner side of the upper teeth and partly the alveolar ridge. Thus, despite the fact that they are traditionally referred to as dental (Stanojević & Popović 2000: 33; Subotić 2005: 65), they will be referred to as alveolar dental stops from this point on. The same term is used by Bakran (1996: 58) in the articulatory description of Croatian /t d/.
duration of the VOT is 53.58 ms, whereas its relation to the place of articulation and vocalic environment adheres to the previously discussed pattern (Table 4).

**Figure 3.** Charts showing the average VOT values (ms) for Serbian and English word-initial stops in relation to the place of articulation and voicing.
The comparison of the results obtained in the analysis of Serbian and English stops suggests that the duration of the voiced interval characterizing phonologically voiced stops and the one of aspiration characterizing their voiceless counterparts tend to be in inverse proportion. Due to the fact that both features can be expressed in terms of the VOT, it follows that long negative VOT of /b d g/ is compensated by relatively short positive VOT of /p t k/, whereas short or zero VOT of /b d g/ correlates with significantly longer positive VOT of /p t k/. While Serbian stops behave with respect to the former pattern, their English counterparts adhere to the latter one (Figure 3).

4. The results of the research: VOT and transfer

Native speakers of Serbian tend to produce English word-initial /b d g/ with considerable vocal fold activity, which points to the presence of negative transfer. According to the conducted measurements, the average duration of the voiced interval for all five subjects is 99.17 ms, which is only 14% shorter in comparison to the average values obtained in the analysis of the corresponding Serbian counterparts. As in the articulation of Serbian voiced stops, the voiced interval increases from /b/ and /d/ to /g/, whereas each stop tends to be produced with slightly more vocal fold activity in front of a high front vowel (Table 5). What is equally important, the ratio between the voiced intervals characterizing the adjacent places of articulation remains rather similar to the one observed in Serbian (cf. Table 2 and Table 5). Accordingly, the influence of Serbian on the production of English /b d g/ can be described as complete. It is reflected in both duration of the voiced phonation and the ratio between the negative VOT values characterizing the adjacent places of articulation.

Table 5. The average duration of the voiced interval (ms) for English word-initial /b d g/ produced by the Serbian subjects

<table>
<thead>
<tr>
<th></th>
<th>/b/</th>
<th>/d/</th>
<th>/g/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>103.2</td>
<td>106.5</td>
<td>116.2</td>
</tr>
<tr>
<td>/i:/</td>
<td>77.6</td>
<td>93.8</td>
<td>97.7</td>
</tr>
<tr>
<td>/i:/</td>
<td>90.4</td>
<td>100.15</td>
<td>106.95</td>
</tr>
<tr>
<td>/i:/</td>
<td>99.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Considering the production of individual subjects, the degree of transfer tends to be rather idiosyncratic. One subject exhibits almost identical values as in the production of Serbian /b d g/, whereas the other four subjects produce English voiced stops with approximately 15-22%
shorter voiced intervals in comparison to the corresponding Serbian equivalents (Table 6).

**Table 6.** The average values of the voiced intervals (ms) for Serbian and English word-initial /b d g/ produced by five Serbian subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Serbian /b d g/</th>
<th>English /b d g/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>133</td>
<td>132.92</td>
</tr>
<tr>
<td>Subject 2</td>
<td>109.08</td>
<td>91.75</td>
</tr>
<tr>
<td>Subject 3</td>
<td>118.42</td>
<td>92.17</td>
</tr>
<tr>
<td>Subject 4</td>
<td>95.75</td>
<td>77.08</td>
</tr>
<tr>
<td>Subject 5</td>
<td>119.92</td>
<td>101.92</td>
</tr>
</tbody>
</table>

**Figure 4.** Chart showing the average values of the voiced intervals (ms) for Serbian and English word-initial /b d g/ produced by five Serbian subjects

While the instances of reducing the vocal fold activity may indicate that Serbian native speakers do perceive the differences between the phonologically voiced stops in the languages in question, the fact that this reduction is not significant as well as the total absence of complete devoicing (Figure 4) undoubtedly point to the strong influence of Serbian.

Interestingly enough, the recorded Serbian subjects also produce English word-initial voiceless stops with considerable aspiration, pointing to the absence of the expected transfer. The average duration of the VOT for all five subjects is 58.37 ms, which is approximately 82% longer in comparison to the average VOT value of the corresponding Serbian equivalents. As expected, the VOT increases from /p/ and /t/ to /k/, while it tends to be slightly longer in front of /i:/ than in front of /o/ (Table 7).
Table 7. The average duration of the VOT (ms) for English word-initial /p t k/ produced by the Serbian subjects

<table>
<thead>
<tr>
<th></th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>44.2</td>
<td>65.9</td>
<td>94.5</td>
</tr>
<tr>
<td>/i:/</td>
<td>31.1</td>
<td>43.1</td>
<td>71.4</td>
</tr>
<tr>
<td></td>
<td>37.65</td>
<td>54.5</td>
<td>82.95</td>
</tr>
</tbody>
</table>

Table 8. The average values of the VOT (ms) for Serbian and English word-initial /p t k/ produced by the Serbian subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Serbian /p t k/</th>
<th>English /p t k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26.92</td>
<td>51.42</td>
</tr>
<tr>
<td>2</td>
<td>34.42</td>
<td>64.08</td>
</tr>
<tr>
<td>3</td>
<td>31.42</td>
<td>48.53</td>
</tr>
<tr>
<td>4</td>
<td>37</td>
<td>53.08</td>
</tr>
<tr>
<td>5</td>
<td>30.5</td>
<td>74.67</td>
</tr>
</tbody>
</table>

Considering the production of individual subjects, the interval of aspiration characterizing English word-initial /p t k/ tends to be 43-145% longer in comparison to the VOT of the corresponding Serbian equivalents (Table 8). Due to the fact that none of the recorded subjects has produced Serbian and English voiceless stops with the same or at least similar VOT values, it is possible to assume that Serbian native speakers have rather good perception of aspiration, causing the discrimination between Serbian and English voiceless stops to be much more successful in comparison to the discrimination between the corresponding voiced counterparts (cf. Figure 4 and Figure 5). This observation is also supported by the fact that two out of five Serbian subjects tend to aspirate English /p t k/ more than the English subjects themselves, which points to the presence of phonetic hypercorrection. The phonetic hypercorrection is especially obvious in the production of the 5th subject whose VOT is approximately 40% longer in comparison to the average VOT measured in the analysis of the English voiceless stops produced by the English native speakers.
Figure 5. Chart showing the average VOT values (ms) for Serbian and English word-initial /p t k/ produced by five Serbian subjects

Still, although the native speakers of Serbian do not produce English voiceless stops with the VOT values typical of their L1, the ratio between the adjacent places of articulation appears to be transferred from Serbian.
Namely, Serbian native speakers tend to produce English /k/ with approximately 52% more aspiration in comparison to English /t/ despite the fact that the average VOT of English /k/ is only 24% longer in comparison to the one of /t/. This ratio appears to be mapped from Serbian whose /k/ is characterized by approximately 48% longer aspiration in comparison to /t/ (Figure 6). As for the ratio between /p/ and /t/, the two languages do not show significant differences and accordingly rather similar ratio is observed in the production of English /p t/ by the Serbian subjects.

Lastly, it should be pointed out that the production of English /p t k/ and /b d g/ by the Serbian speakers does not exhibit the relation of inverse proportion in the VOT values. In other words, those Serbian subjects who produce English voiceless stops with considerable aspiration will not produce the corresponding voiced counterparts with significantly lesser vocal fold activity and vice versa (Figure 7). Thus, even on the assumption that Serbian native speakers do have rather good perception of aspiration characterizing English voiceless stops, this perception appears to be artificial, possibly resulting from the intensive phonetic training. As for the production of the corresponding Serbian counterparts, the inverse proportion between the VOT values of /p t k/ and /b d g/ is retained (Figure 8).

**Figure 7.** Chart showing the average duration of the VOT (ms) for English word-initial stops produced by five Serbian subjects
5. Concluding remarks

The results of the research show that Serbian and English rely on different VOT values in order to make the sets of phonologically voiced and voiceless stops as distinct as possible. While the articulation of Serbian word-initial voiced stops involves considerable vocal fold activity, resulting in long negative VOT, the corresponding English equivalents tend to be completely devoiced in this position, resulting in zero VOT. Accordingly, the voiceless stops of Serbian are hardly audibly aspirated, whereas the production of their English counterparts involves considerable aspiration.

Equally importantly, Serbian native speakers show heavy reliance on their L1 experience in the production of English stop consonants. The transfer is particularly present in the articulation of /b d g/ which are produced with considerable phonation without a single instance of complete devoicing. In addition, the ratio between the voiced intervals of the adjacent places of articulation is rather similar to the one observed in Serbian. The production of English /p t k/ is characterized by significantly longer VOT values in comparison to Serbian, which together with the occasional presence of phonetic hypercorrection suggests the absence of the expected transfer. The absence of transfer may result from either good
perceptual abilities of the Serbian subjects or possibility that the strong presence of aspiration of English /p t k/ was more stressed and practised in comparison to the voicing properties of English /b d g/ within the phonetic training. In order to find an answer to the question, it would be necessary to analyze the speech of those Serbian native speakers who have not received any phonetic training in English pronunciation. Still, whatever the reason is, the influence of Serbian on the production of English voiceless stops is not completely absent. It can be observed in the act of mapping the VOT ratios between the adjacent places of articulation from Serbian into English.

References

The Evolution of a Phonetic Phenomenon: The Case of Voice Onset Time in Serbian Intermediate EFL Learners

Biljana Ćubrović

Outline

This paper looks into the nature of the production and acquisition of aspirated voiceless stops /p t k/ of advanced EFL speakers of Serbian language background, whose language clearly lacks them. A list of selected monosyllabic English words was recorded and compared to their closest approximations in Serbian. The informants, 2 male and 2 female speakers are prompted to read the selected tokens in carrier sentences. Acoustic analysis is carried out so as to show to what extent Serbian and English differ in the production of voiceless stops. The analysis also throws some new light on the acquisition of aspiration in Serbian native speakers in their English pronunciation.

1. Introduction

In order to define the interval between the onset of laryngeal vibration and the release of a stop, the term VOT is used extensively in the phonetic literature. Some languages seem to make more use of this parameter, where others do not. Starting from this assumption, Lisker and Abramson (1964) look into the VOT values for 11 different languages and set up a scale which has come to be used in later research on stop consonants. To sum their research up, Abramson (1977: 296) later pointed out that they

---

1 This chapter is part of the project no. 178019 supported by the Serbian Ministry of Education and Science.
adopted “the convention of assigning a timing value of zero to the moment of stop release, negative values to voicing lead, and positive values to voicing lag”. What they found was “a tri-modal distribution of VOT values” (ibid: 296). The first category centers at -100 ms for a range of values by which voiced unaspirated stops are represented. The second mode amounts to +10ms and corresponds to what we know as voiceless unaspirated stops. The third category of stops centers at +75ms and corresponds to voiceless aspirated stops.

Further research relies heavily on the results that Abramson and Lisker (1964) obtained in their cross-language study, but it was only a trigger for numerous studies in which VOT was measured in normal adult and children’s speech, and also in a number of speech disorders both in adults and children. Kent & Read (1996: 108) assign similar values of VOT for stops, and point out that the delay in voiceless consonants is on the order of 25-100 ms. Kent & Read (1996) truly emphasize that there can hardly be deduced a single value of VOT that could be used by all speakers or across phonetic contexts. These values are in line with Abramson and Lisker’s (1964) numerical data.

VOT also depends on the place of articulation. Whalen et al. (2007: 342) claim that VOT values of labial stops are consistently shorter than the values for lingual stops. However, it seems that some speaker variation exists, and some studies claim that alveolar stops have a shorter VOT values than velars (Lisker & Abramson 1967; Zue 1976; Weismer 1979), whereas others claim that these values are the same (Docherty 1989; Cooper 1991). Such differences may be accounted for by the speech rate at which speakers enunciate the analyzed material. More often than not, the details of the recording procedure are not clearly stated, so tempo as a factor in VOT values has to be considered. Aerodynamics is another factor that explains the differences in VOTs across various places of articulation: the lips are extremely movable due to their small mass, as compared to the soft palate which is not as agile. Velar contact is slow and so-called “double burst” (even triple burst) may ensue in the articulation of /k g/.

Serbian and English show similarities in the places of articulation for bilabial /p b/ and velar /k g/, but the English alveolar consonants /t d/ have dental counterparts in Serbian. This variation should be borne in mind when analyzing VOT values in these two languages.

2. Experiment design

A list of 27 English and Serbian words, monosyllables or disyllables, was recorded. Wherever possible, minimal or near minimal pairs, were
used in order to neutralize the potential differences which could have been created by deviations in the phonetic environments in the English and Serbian tokens. Nine vowels, both short and long in duration, were analyzed in word-initial, accented positions. Word-initial positions are the standard positions that are considered when VOTs are measured in a single language or cross-linguistically. All vowels under scrutiny were preceded by one of the voiceless plosives /p t k/. The selection of 27 phonetic contexts provides a common vocalic denominator typical of English and Serbian. The English vowel qualities that are studied are /e i: i u: u ɔ ɔ: aː/. Their Serbian phonetic approximations are /i u o a/, and the phonetic contexts with both long and short pitch accents in Serbian are taken into account. Additionally, the short Serbian /e/, is looked at, whereas its long Serbian counterpart /eː/ is disregarded, e.g. têma (Eng. topic), pêtak (Eng. Friday), due to the lack of this vowel quality in English. The rationale lying behind the study is that the phonetic conditions are as similar as possible in the two apparently different languages. The English vowel inventory is much more diversified compared to Serbian, which has a simple five-vowel system /a e i o u/, but a rather complex pitch accent system with four types of pitch accents (short falling, short rising, long falling, and long rising).

Each token was recorded three times in carrier sentences. Carrier sentences are used so as to resemble natural speech (prompted speech recorded for the purpose of phonetic analysis is usually artificial and also slower in tempo). All tokens were placed in accented positions and informants were instructed to stress them. Two female and two male Serbian speakers are all speakers of English at an (upper)-intermediate level. They were all freshmen at the English Department, and were twenty years old at the time of the recording. Their mean age was 20:3. All four speakers were born in Belgrade and have not visited the English speaking territory.

Recordings were made in Praat, version 5.1.33, at a sampling rate of 22,050 Hz, using a Sennheiser PC156 noise cancelling microphone. Recordings were analysed in the same software package, with the help of spectrograms and waveforms.

3. Results

The following abbreviations are used for the four informants: M1 (male speaker no. 1), M2 (male speaker no. 2), F1 (female speaker no. 1), and F2 (female speaker no. 2). The main hypothesis postulated before the experiment is that VOT values are shorter for Serbian tokens than for
Serbian English tokens, due to the fact that Serbian does not recognize aspiration as a distinctive feature of Serbian stops. Ranges of VOT values are given first for each individual speaker, followed by mean values for each CV sequence (presented in graphs underneath).

M1 VOT average\(^2\) value for the Serbian tokens containing /p/ range from 13 ms for /pe-/ and 39.67 ms for /pu-/ . The average VOTs for the Serbian tokens containing /t/ start from 23 ms for /te/ and are the highest for /ti/ , 53.33 ms. The velar plosive /k/ exhibits highest values of all three consonants, that are lowest for /ki/ , 34 ms and highest for /ku/ , 54 ms. The Serbian English VOTs were expected to be lower than those shown for Serbian, but the data offer some unexpected results. Namely, M1’s VOT values are not consistently higher for Serbian English, though intermediate speakers had been familiarized with the important role of aspiration in English before the recording took place. The following phonetic contexts have lower VOT values in Serbian English than in Serbian: /pu:/ (39.67 ms vs. 32.00 ms), /p<sub>1</sub>-/ (35 ms vs. 23 ms), /p<sub>3</sub>-/ (27.33 ms vs. 18.67 ms), /t<sub>1</sub>-/ (31.33 ms vs. 24.67 ms), /t<sub>7</sub>-/ (32.67 ms vs. 22.33 ms), /k<sub>1</sub>-/ (34 ms vs. 31.67 ms), /k<sub>7</sub>-/ (39 ms vs. 37.67 ms), and /k<sub>8</sub>-/ (49.67 ms vs. 49 ms).

The VOT measurements are given in Graph 1 below for the first male speaker. VOT values for Serbian tokens are given in the first column (ms), and these are followed by the values for Serbian English tokens in column 2.

M2 VOT average value for the Serbian tokens containing /p/ range from 14.33 ms for /p<sub>3</sub>-/ and 50.67 ms for /pu:-/. The average VOTs for the Serbian tokens containing /t/ start from 20.67 ms for /ta-/ and are the highest for /tu:-/, 47.67 ms. The velar plosive /k/ performs highest values of all three consonants, that are lowest for /ka:-/, 46.67 ms and highest for /ki:-/, 80.67 ms. The hypothesis that the Serbian English VOTs are lower than those shown for Serbian again proved somewhat unstable. M2’s VOT values are not invariably higher for Serbian English. The following phonetic environments exhibit lower VOT values in Serbian English than in Serbian: /pe-/ (14.67 ms vs. 14.33 ms), /ki-/ (69.33 ms vs. 63.67 ms), /k<sub>0</sub>-/ (56.33 ms vs. 55 ms), and /k<sub>a</sub>-/ (60 ms vs. 57 ms). It is evident that M2’s unexpected oscillations in VOTs between Serbian and Serbian English are usually very slight.

\(^2\) By *average* I mean average values for the three occurrences of a single token.
The VOT measurements are given in Graph 2 below for the second male speaker. VOT values for Serbian tokens are given in the first column (ms), and these are followed by the values for Serbian English tokens in column 2.

**Graph 1.** Mean VOT values for M1 speaker

**Graph 2.** Mean VOT values for M2 speaker

F1 VOT average value for the Serbian tokens containing /p/ range from 15.33 ms for /pi-/ and 44.67 ms for /pu:/. The average VOTs for the Serbian tokens containing /t/ start from 20 ms for /ta-/ and are the highest for /ti:-/, 35.67 ms. The velar plosive /k/ exhibits highest values of all three consonants, and the measurements are lowest for /ku-/, 54 ms and highest
for /ki:–/, 90.33 ms. In the case of F2, the Serbian English VOTs are consistently compared to those characteristic of Serbian tokens, which is in line with the original expectations.

The VOT measurements are given in Graph 3 below for the first female speaker. VOT values for Serbian tokens are given in the first column (ms), and these are followed by the values for Serbian English tokens in column 2.

F2 VOT average value for the Serbian tokens containing /p/ range from 9.67 ms for /pi–/ and 29.67 ms for /po–/. The average VOTs for the Serbian tokens containing /t/ start from 14.33 ms for /te–/ and are the highest for /to–/, 23.67 ms. The velar plosive /k/ exhibits highest values of all three consonants, and the measurements are lowest for /ka–/, 30 ms and highest for /ki–/, 46.33 ms. In the case of F2, the Serbian English VOTs are consistently and sometimes significantly higher than those for Serbian tokens, which conforms to the original hypothesis.

The VOT measurements are given in Graph 4 below for the second female speaker. VOT values for Serbian tokens are given in the first column (ms), and these are followed by the values for Serbian English tokens in column 2.
4. Discussion and final remarks

One of the most striking findings that resulted from the acoustic measurements is a tendency of two male speakers to use less aspiration in their English pronunciation than in Serbian, where it is not distinctive. This may be accounted for by the lower level of EFL proficiency in the two male speakers. Such articulatory qualities of some aspirated stops/sequences of stops and vowels may be perceived as purely idiosyncratic. In order to reach some more general conclusions, average VOT values for all four speakers are analyzed, as shown in the tables to follow.

### Table 1: VOT values for Serbian stops

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Range</th>
<th>No. of tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>22.06</td>
<td>13:36</td>
<td>216</td>
</tr>
<tr>
<td>/t/</td>
<td>27.62</td>
<td>21:39</td>
<td>216</td>
</tr>
<tr>
<td>/k/</td>
<td>51.69</td>
<td>42:64</td>
<td>216</td>
</tr>
</tbody>
</table>
Table 2: VOT values for Serbian English stops

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Range</th>
<th>No. of tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>31.44</td>
<td>19:46</td>
<td>216</td>
</tr>
<tr>
<td>/t/</td>
<td>48.85</td>
<td>34:65</td>
<td>216</td>
</tr>
<tr>
<td>/k/</td>
<td>62.53</td>
<td>54:76</td>
<td>216</td>
</tr>
</tbody>
</table>

The average VOT values are as predicted. Cumulative values of Serbian stops seem lower than those of Serbian English stops. However, individual deviations in VOTs, where Serbian stops had more aspiration compared to Serbian English stops have been annihilated when average measurements are generated.

What is easily noticed in Tables 1 and 2 is a systematically arranged increase in VOT values in Serbian English stops, starting from the bilabial /p/, to the alveolar /t/, ending in the velar /k/. Similar results were obtained in an earlier mirror study conducted with proficient speakers of Serbian English. The average VOT values for this group of speakers for Serbian English were slightly higher (39 ms for /p/, 64 ms for /t/, and 71 ms for /k/) (Čubrović 2011: 15). This may lead us to a conclusion that aspiration is a phonetic phenomenon that is acquired gradually. Non-native speakers of English are obviously able to slowly stretch their VOT, until its numerical values reach the point where native speakers perceive the relevant consonants as voiceless aspirated.

These conclusions may only be considered tentative due to a relatively small sample used. A more comprehensive longitudinal study with a more significant, representative sample is required to confirm the findings obtained in this small-scale research study.

References


Outline

The aim of this paper is to start supplementing current descriptions of connected speech in English, which do not seem to have been accounted for satisfactorily due to the phonetics’ past focus on relatively formal style. As a first step, I revise the descriptions of weak forms of grammatical and other common words. The source of the data is the Buckeye Corpus of Conversational Speech (Pitt et al. 2007). Using the “weakners” list in Obendorfer (1998) as the starting point, I first searched for words which seemed to be missing in Obendorfer’s list, and suggested some possible additions. Then actual pronunciations of selected words in the list were searched, and a large number of undescribed weak forms were found. The collapsing of the different forms into a much smaller, manageable number of groups is obviously necessary, and can lead to a new understanding of phonological processes.¹

1. Background and purpose

Phonetics has traditionally been focused on relatively formal speech styles. The fact that “articulatory reduction and elaboration in speech” and “spontaneous speech” were predicted to “dominate future phonetic research” (Kohler 2000) underlines this tendency. Also, studies of small

¹ This paper is based on Makino (2009, 2010). The research for those studies was supported by Grant-in-Aid for Scientific Research (B) No. 20300265 (project leader: Nobuaki Minematsu) from the Japan Society for the Promotion of Science, and a Chuo University Grant for Special Research.
datasets, which are observed in situ or artificially collected in laboratory settings, have been the norm.

Technological advances in computation and mass storage are promising a shift to the analysis of corpora of natural speech that are thousands of times larger. The publication of the Buckeye Corpus of Conversational Speech (Pitt et al. 2007), a phonetically transcribed spoken corpus of American English, is one of the major developments in this trend in the field.

The purpose of this paper is to start supplementing current descriptions of connected speech in English, and the area I will first deal with is “weak forms” of grammatical and other common words (or “function words” in short). The standard textbook accounts (e.g., Kenyon 1997; Cruttenden 2008) do not show all the pronunciations of what we come across in everyday speech in English. This may be due to their considerations of foreign learners as well as the above-mentioned focus on formal speech. But everyday reality should not be ignored, and accounts of more informal style, or casual style are necessary.

The Buckeye Corpus, which is based on casual interviews, is a welcome addition which can be used as a rich data source for conversational speech. People who developed the corpus had different things in mind, however, and their publications do not include descriptions of English (cf. Ohio State University Department of Psychology 2007, and Johnson 2004, 2007). Thus I decided to take advantage of this rich resource, no matter how elementary this kind of study may sound.

This study will be a first step toward a more satisfactory and useful description of the pronunciation of English.

2. The current list of weak forms

Obendorfer (1998) lists some 100 words which are weakened in proper (phonetic) environments (called “weakeners”) and their weak form pronunciations. To my knowledge, it is the most comprehensive list of its kind and should be the starting point of the present study.

The definition of the term “weak form” which I adopt is also essentially from Obendorfer (1998: 28):

(1) [W]eak forms are paradigmatically non-basic phonological word forms, or parts of word forms.
(2) [T]hey represent morphosyntactic words in certain non-prominent contexts.
(3) [T]heir phonological shape is semi-reduced, reduced or cliticized.
(4) [T]hey are the products of an ultimately idiosyncratic process.

Definition clause (4) is the most important and gives me justification for conducting the study of weak forms as the study of words.

Table 1 shows the list of weakeners and their weak forms given in Obendorfer (1998: 206-10). The pronunciations given in the original table are given in British RP, but I have adapted them to American English forms in order to facilitate the comparison with the forms appearing in the data source used in this study (see §3 below).

I have left out from Obendorfer’s list the archaic forms of verbs in the second person singular conjugation, which are unlikely to be used in spontaneous speech. I also have not included words in his “polysyllabic” list and “chained weakeners” (i.e., contracted words) list.

In an attempt to make the table a more comprehensive reflection of the current description, I have added weak forms found in Wells (2008; LPD3) and Roach, Hartman and Setter (2006; EPD17) (in the right two columns). LPD3 listed two words not listed by Obendorfer, which appear in the last part of the table.

**Table 1.** The current list of weakeners and their weak forms

<table>
<thead>
<tr>
<th>Part of speech</th>
<th>Obendorfer</th>
<th>LPD3</th>
<th>EPD17</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>det</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>an</td>
<td>det</td>
<td>øn</td>
<td>øn</td>
</tr>
<tr>
<td>the</td>
<td>det</td>
<td>öi, öo</td>
<td>öi, öo</td>
</tr>
<tr>
<td>be</td>
<td>aux, verb</td>
<td>bi</td>
<td>bi, bo</td>
</tr>
<tr>
<td>been</td>
<td>aux, verb</td>
<td>bin</td>
<td>bin</td>
</tr>
<tr>
<td>am</td>
<td>aux, verb</td>
<td>øm, m</td>
<td>øm, m</td>
</tr>
<tr>
<td>are</td>
<td>aux, verb</td>
<td>ø</td>
<td>ø, r</td>
</tr>
<tr>
<td>is</td>
<td>aux, verb</td>
<td>z, s</td>
<td>z, s</td>
</tr>
<tr>
<td>was</td>
<td>aux, verb</td>
<td>wəz, wz</td>
<td>wəz</td>
</tr>
<tr>
<td>were</td>
<td>aux, verb</td>
<td>wə</td>
<td>wə</td>
</tr>
<tr>
<td>had</td>
<td>aux, verb</td>
<td>həd, əd, d</td>
<td>həd, əd, d</td>
</tr>
<tr>
<td>has</td>
<td>aux, verb</td>
<td>həz, əz, z, s</td>
<td>həz, əz, z, s</td>
</tr>
<tr>
<td>have</td>
<td>aux, verb</td>
<td>həv, əv, v</td>
<td>həv, əv, v</td>
</tr>
<tr>
<td>did</td>
<td>aux, verb</td>
<td>dəd</td>
<td>səd, d</td>
</tr>
<tr>
<td>do</td>
<td>aux, verb</td>
<td>də, də, d</td>
<td>də, də, d</td>
</tr>
<tr>
<td>does</td>
<td>aux, verb</td>
<td>dəz</td>
<td>dəz, dz</td>
</tr>
<tr>
<td>can</td>
<td>aux</td>
<td>kən</td>
<td>kən, kən</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>could</td>
<td>aux</td>
<td>kəd, kd</td>
<td>kəd</td>
<td>kəd</td>
<td></td>
</tr>
<tr>
<td>may</td>
<td>aux</td>
<td>mi, mə</td>
<td>mə</td>
<td>mə</td>
<td></td>
</tr>
<tr>
<td>must</td>
<td>aux</td>
<td>mast, mst</td>
<td>məst, mas</td>
<td>məst, mas</td>
<td></td>
</tr>
<tr>
<td>shall</td>
<td>aux</td>
<td>jəl, jəl, jə, jə</td>
<td>jəl, jə</td>
<td>jəl, jə</td>
<td></td>
</tr>
<tr>
<td>should</td>
<td>aux</td>
<td>jəd, jəd</td>
<td>jəd, jəd, jət</td>
<td>jəd, jəd, jət</td>
<td></td>
</tr>
<tr>
<td>will</td>
<td>aux</td>
<td>wəl, wəl, l</td>
<td>wəl, wəl, wəl, l</td>
<td>wəl, wəl, wəl, l</td>
<td></td>
</tr>
<tr>
<td>would</td>
<td>aux</td>
<td>wəd, wəd, d</td>
<td>wəd, wəd</td>
<td>wəd, wəd, d</td>
<td></td>
</tr>
<tr>
<td>at</td>
<td>prep</td>
<td>ət</td>
<td>ət</td>
<td>ət</td>
<td></td>
</tr>
<tr>
<td>by</td>
<td>prep</td>
<td>bə, bə</td>
<td>bə, bə</td>
<td>bə, bə</td>
<td></td>
</tr>
<tr>
<td>for</td>
<td>prep</td>
<td>fə, fr</td>
<td>fə, fr</td>
<td>fə, fr</td>
<td></td>
</tr>
<tr>
<td>from</td>
<td>prep</td>
<td>frəm</td>
<td>frəm</td>
<td>frəm</td>
<td>frəm</td>
</tr>
<tr>
<td>in</td>
<td>prep</td>
<td>ən</td>
<td>ən</td>
<td>ən</td>
<td></td>
</tr>
<tr>
<td>of</td>
<td>prep</td>
<td>əv, ə</td>
<td>əv, ə</td>
<td>əv</td>
<td></td>
</tr>
<tr>
<td>on</td>
<td>prep</td>
<td>ən</td>
<td>ən</td>
<td>ən</td>
<td></td>
</tr>
<tr>
<td>per</td>
<td>prep</td>
<td>pə</td>
<td>pə</td>
<td>pə</td>
<td></td>
</tr>
<tr>
<td>to</td>
<td>prep</td>
<td>tu, tə</td>
<td>tu, tə</td>
<td>tu, tə</td>
<td>tə, tə, tu</td>
</tr>
<tr>
<td>up</td>
<td>adv, prep</td>
<td>əp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with</td>
<td>prep</td>
<td>wəd, wəd, wəd</td>
<td>wəd, wəd, wəd</td>
<td>wəd, wəd, wəd</td>
<td></td>
</tr>
<tr>
<td>and</td>
<td>conj</td>
<td>ənd, n</td>
<td>ənd, ən, nd, n, m, n, n, m, n, n,</td>
<td>ənd, ən, nd, n, m, n, n, m, n, n, m, n, n,</td>
<td></td>
</tr>
<tr>
<td>but</td>
<td>conj</td>
<td>bət, bt</td>
<td>bət</td>
<td>bət</td>
<td></td>
</tr>
<tr>
<td>nor</td>
<td>conj</td>
<td>nə</td>
<td>nə</td>
<td>nə</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td>conj</td>
<td>ə</td>
<td>ə</td>
<td>ə</td>
<td></td>
</tr>
<tr>
<td>so</td>
<td>adv, conj</td>
<td>sə</td>
<td>sə</td>
<td>sə</td>
<td></td>
</tr>
<tr>
<td>as</td>
<td>adv, conj, prep</td>
<td>əz</td>
<td>əz</td>
<td>əz</td>
<td></td>
</tr>
<tr>
<td>if</td>
<td>conj</td>
<td>əf</td>
<td>əf</td>
<td></td>
<td></td>
</tr>
<tr>
<td>than</td>
<td>conj</td>
<td>dən, dən, dən</td>
<td>dən, dən, dən</td>
<td>dən, dən, dən</td>
<td></td>
</tr>
<tr>
<td>till</td>
<td>conj, prep</td>
<td>təl</td>
<td>təl, təl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>what</td>
<td>rel</td>
<td>wət</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>when</td>
<td>rel</td>
<td>wən</td>
<td>wən</td>
<td></td>
<td></td>
</tr>
<tr>
<td>who</td>
<td>rel</td>
<td>u, u</td>
<td>hu, u</td>
<td></td>
<td>hu, u</td>
</tr>
<tr>
<td>whom</td>
<td>rel</td>
<td>um, huz, uz</td>
<td>hum, um</td>
<td></td>
<td></td>
</tr>
<tr>
<td>whose</td>
<td>rel</td>
<td>um, hum, um</td>
<td>huz, uz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>that</td>
<td>conj, det, pron, rel</td>
<td>dət</td>
<td>dət</td>
<td>dət, dət</td>
<td></td>
</tr>
<tr>
<td>he</td>
<td>pron</td>
<td>hi, i, i</td>
<td>hi, i</td>
<td>hi, i</td>
<td></td>
</tr>
<tr>
<td>her</td>
<td>pron (det)</td>
<td>hə, ə, ɐ</td>
<td>hə, ə, ɐ</td>
<td>hə, ə</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>him</td>
<td>pron</td>
<td>im</td>
<td>him, im</td>
<td>ʃhəm, ʃəm</td>
<td>him, im</td>
</tr>
<tr>
<td>his</td>
<td>pron (det)</td>
<td>ɨz</td>
<td>hiz, iz, ʃhəz, ʃəz</td>
<td>ɨz</td>
<td></td>
</tr>
<tr>
<td>it</td>
<td>pron</td>
<td>ət, t</td>
<td>ʃət</td>
<td></td>
<td></td>
</tr>
<tr>
<td>me</td>
<td>pron</td>
<td>mi</td>
<td>mi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>my</td>
<td>pron (det)</td>
<td>mi, mə</td>
<td>mi, mə</td>
<td>mi</td>
<td></td>
</tr>
<tr>
<td>she</td>
<td>pron</td>
<td>ʃi</td>
<td>ʃi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>their</td>
<td>pron (det)</td>
<td>ʃə</td>
<td>ʃə</td>
<td>ʃə</td>
<td></td>
</tr>
<tr>
<td>them</td>
<td>pron</td>
<td>ʃəm</td>
<td>ʃəm, ʃm</td>
<td>ʃəm, ʃm, ʃm, m</td>
<td></td>
</tr>
<tr>
<td>they</td>
<td>pron</td>
<td>ʃəm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>us</td>
<td>pron</td>
<td>əs</td>
<td>əs, ʃəz</td>
<td>əs, s</td>
<td></td>
</tr>
<tr>
<td>we</td>
<td>pron</td>
<td>wi</td>
<td>wi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>you</td>
<td>pron</td>
<td>ju, ʃə</td>
<td>ju, ʃə</td>
<td>ju, ʃə</td>
<td></td>
</tr>
<tr>
<td>your</td>
<td>pron (det)</td>
<td>ʃə</td>
<td>ʃə</td>
<td>ʃə</td>
<td></td>
</tr>
<tr>
<td>one</td>
<td>pron</td>
<td>ʃəm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>any</td>
<td>det</td>
<td>ʃəni</td>
<td>ʃəni, ni</td>
<td>ʃəni, ni</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>det</td>
<td>ʃən</td>
<td>ʃən</td>
<td></td>
<td></td>
</tr>
<tr>
<td>some</td>
<td>det</td>
<td>ʃəm, sm</td>
<td>ʃəm, sm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>such</td>
<td>det</td>
<td>ʃəf</td>
<td>ʃəf</td>
<td></td>
<td></td>
</tr>
<tr>
<td>this</td>
<td>det, pron</td>
<td>ʃəs, ʃəs</td>
<td>ʃəs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sure</td>
<td>adj</td>
<td>ʃə</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>just</td>
<td>adv</td>
<td>ʃə nast</td>
<td>ʃə nast, ʃə nastst</td>
<td>ʃə nastst</td>
<td></td>
</tr>
<tr>
<td>not</td>
<td>adv</td>
<td>ʃət</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>then</td>
<td>adv</td>
<td>ʃən</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>there</td>
<td>adv</td>
<td>ʃə, ʃə</td>
<td>ʃə</td>
<td></td>
<td></td>
</tr>
<tr>
<td>come</td>
<td>verb</td>
<td>ʃəm</td>
<td>ʃəm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>get</td>
<td>verb</td>
<td>ʃət</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>go</td>
<td>verb</td>
<td>ɡə, ɡə</td>
<td>ʃɡə, ʃɡu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>said</td>
<td>verb</td>
<td>ʃəd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>says</td>
<td>verb</td>
<td>ʃəz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sit</td>
<td>verb</td>
<td>ʃət</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thank</td>
<td>verb</td>
<td>(ŋ)k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ma'am</td>
<td>noun</td>
<td>ʃəm</td>
<td>ʃəm, əm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saint</td>
<td>noun</td>
<td>ʃənt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sir</td>
<td>noun</td>
<td>ʃə</td>
<td>ʃə</td>
<td>ʃə</td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>noun</td>
<td>ʃəm</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Method

As mentioned above, the data source in this study is the Buckeye Corpus of Conversational Speech, compiled in the Department of Psychology at the Ohio State University. This is a phonetically transcribed corpus of informal interviews (total 300,000 words) of 40 speakers of English from the Columbus, Ohio area.

According to the corpus manual, the transcription procedure was:
1. Writing down orthographic words by using Soundscriber software.
2. Automatically generating phonetic transcriptions and aligning the word transcription and the phonetic transcription to the media file, by using ESPS Aligner software.
3. Manually correcting the transcriptions generated by Aligner.

There are some limitations in the corpus. It does not transcribe word stress, sentence accent or intonation. Thus, it is impossible to decide whether the segments iy (iː), ih (iː), uh (uː), ah (oː) and er (ɛr) correspond to strong or weak vowels just by looking at the transcription. It is possible if you listen to the media, but that has been impractical, given the large amount of data.

I used the SpeechSearcher software which accompanies the corpus for data search and retrieval. Regarding the above difficulty with /iː, i, uː, o, ə-, ə-ə/, I assumed that if these vowels in the actual phone tier corresponded to the different symbols in the dictionary phone tier (i.e., citation forms), they were weak vowels. For example, [ʊ] is strong if the citation form is also /ʊ/ but weak if not. For which, whose citation form contains /ɪ/, [wɪtʃ] is strong and [wʊtʃ] is weak.

4. Possible revision of the list: Words

Assuming that the same kind of words (of the same grammatical category) behave the same way in terms of accentability, I find the following possible gaps in the current weakener list:

<table>
<thead>
<tr>
<th>times</th>
<th>noun</th>
<th>тəmз</th>
</tr>
</thead>
<tbody>
<tr>
<td>well</td>
<td>interj.</td>
<td>wəl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wəl, wəl</td>
</tr>
</tbody>
</table>

Only in LPD3:

<table>
<thead>
<tr>
<th>I</th>
<th>pron</th>
<th>ə, a</th>
</tr>
</thead>
<tbody>
<tr>
<td>its</td>
<td>pron</td>
<td>əts</td>
</tr>
</tbody>
</table>
(1) Among the relatives, “how”, “where”, “which” and “why” are not listed.
(2) In the personal pronoun category, “our” is absent.
(3) In the auxiliary verb group, “might” is missing.
(4) Main verbs are, of course, marginal among the weakeners, but I find it inconsistent that some forms in the conjugation paradigm of verbs which are listed are absent: “Said” and “says” are listed but not “say,” “come” is listed but not “came,” “go” is listed but “went” and “gone” are not, and “get” is listed but “gets,” “got” and “gotten” are not.

I will look at the findings for each item in the next section.

5. Findings: Words to be added to the list

Tables 2-7 show the search results for “where,” “gets,” “our,” “most,” “which” and “went” respectively. They all include a considerable number of forms with a reduced vowel or without vowels at all, and the results suggest strongly that they can be added to the list of weakeners.

Table 2. Strong and weak forms of “where”
### Table 3. Strong and weak forms of “gets”

<table>
<thead>
<tr>
<th>Forms</th>
<th>Tokens</th>
<th>Percentage</th>
<th>Strong / weak %</th>
</tr>
</thead>
<tbody>
<tr>
<td>gets</td>
<td>33</td>
<td>42.3</td>
<td>Strong 53.8%</td>
</tr>
<tr>
<td>ges</td>
<td>7</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>gds</td>
<td>1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>gez</td>
<td>1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>gts</td>
<td>29</td>
<td>37.2</td>
<td>Weak 46.2%</td>
</tr>
<tr>
<td>gis</td>
<td>2</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>gitf</td>
<td>1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>gids</td>
<td>1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>gidz</td>
<td>1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>giz</td>
<td>1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>kits</td>
<td>1</td>
<td>1.3</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4. Strong and weak forms of “our”

<table>
<thead>
<tr>
<th>Forms</th>
<th>Tokens</th>
<th>Percentage</th>
<th>Strong / weak %</th>
</tr>
</thead>
<tbody>
<tr>
<td>our</td>
<td>257</td>
<td>73.6</td>
<td>Strong 79.7%</td>
</tr>
<tr>
<td>aur</td>
<td>16</td>
<td>4.6</td>
<td></td>
</tr>
<tr>
<td>a:</td>
<td>3</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>ar</td>
<td>2</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>ər</td>
<td>30</td>
<td>8.6</td>
<td>Weak 18.3%</td>
</tr>
<tr>
<td>ə</td>
<td>29</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>ə</td>
<td>5</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Other forms</td>
<td>7</td>
<td>2.0</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

### Table 5. Strong and weak forms of “most”

<table>
<thead>
<tr>
<th>Forms</th>
<th>Tokens</th>
<th>Percentage</th>
<th>Strong / weak %</th>
</tr>
</thead>
<tbody>
<tr>
<td>most</td>
<td>111</td>
<td>47.0</td>
<td>Strong 85.2%</td>
</tr>
<tr>
<td>mous</td>
<td>87</td>
<td>36.9</td>
<td></td>
</tr>
<tr>
<td>moust</td>
<td>3</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>mast</td>
<td>14</td>
<td>5.9</td>
<td>Weak 11.0%</td>
</tr>
<tr>
<td>mas</td>
<td>12</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>Other forms</td>
<td>9</td>
<td>3.8</td>
<td>3.8%</td>
</tr>
</tbody>
</table>
Table 6. Strong and weak forms of “which”

<table>
<thead>
<tr>
<th>Forms</th>
<th>Tokens</th>
<th>Percentage</th>
<th>Strong / weak %</th>
</tr>
</thead>
<tbody>
<tr>
<td>źwŋ</td>
<td>257</td>
<td>80.6</td>
<td>Strong 84.6%</td>
</tr>
<tr>
<td>hwźwŋ</td>
<td>7</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>źwŋ</td>
<td>6</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>wźf</td>
<td>14</td>
<td>4.4</td>
<td>Weak 9.4%</td>
</tr>
<tr>
<td>wźf</td>
<td>13</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>tʃ</td>
<td>3</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Other forms</td>
<td>19</td>
<td>6.0</td>
<td>6.0%</td>
</tr>
</tbody>
</table>

Table 7. Strong and weak forms of “went”

<table>
<thead>
<tr>
<th>Forms</th>
<th>Tokens</th>
<th>Percentage</th>
<th>Strong / weak %</th>
</tr>
</thead>
<tbody>
<tr>
<td>wən</td>
<td>123</td>
<td>28.8</td>
<td>Strong 87.6%</td>
</tr>
<tr>
<td>went</td>
<td>109</td>
<td>25.5</td>
<td></td>
</tr>
<tr>
<td>wɛ瑞</td>
<td>55</td>
<td>12.9</td>
<td></td>
</tr>
<tr>
<td>wɛʔ</td>
<td>30</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>wɛʔ</td>
<td>17</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>wɛʔ</td>
<td>10</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>wɛt</td>
<td>8</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>wend</td>
<td>6</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>wɛ</td>
<td>5</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>we</td>
<td>5</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>wɛt</td>
<td>4</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>wɛm</td>
<td>2</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>wən</td>
<td>17</td>
<td>4.0</td>
<td>Weak 7.5%</td>
</tr>
<tr>
<td>wɛ瑞</td>
<td>5</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>wɛʔ</td>
<td>5</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>wɛnt</td>
<td>3</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>wɛnt</td>
<td>2</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Other forms</td>
<td>21</td>
<td>4.9</td>
<td>4.9%</td>
</tr>
</tbody>
</table>

The results for the other candidates to the list are:
(1) here = 3.2% weak
(2) why = 2.6% weak
(3) though = 2.6% weak
(4) how = 2.5% weak
It is difficult to draw a line here. Is 3% of the tokens enough to be included in the standard description? I could include these, but I could equally leave out “went” and “which,” which have fewer than 10% weak occurrences.

For the following five items, no tokens of “weak” forms have been found. This does not mean that these items cannot reduce. The absence may just be accidental. This is the limitation of the corpus-based study.

(5) got, gotten, gone, came, might

6. Findings and Possible Revisions: Weak Forms of Individual Words

6.1. An overview

Weak forms of the selected words are given in Table 8. Here, forms not listed in Table 1 are in boldface.

<table>
<thead>
<tr>
<th>Word</th>
<th>Weak forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>ə, i, ı, u, ə, i, u</td>
</tr>
<tr>
<td>an</td>
<td>ən, ən, ən, ən, ən, ən</td>
</tr>
<tr>
<td>been</td>
<td>bɪn, bɪn, bɪn, bɪn, bɪn, bɪn</td>
</tr>
<tr>
<td>was</td>
<td>wəz, wəz, wəz, wəz, wəz, wəz, wəz</td>
</tr>
<tr>
<td>have</td>
<td>əv, ə, iv, i, v, həv</td>
</tr>
<tr>
<td>do</td>
<td>dʊ, dʊ, dʊ, dʊ, dʊ, dʊ, dʊ</td>
</tr>
<tr>
<td>from</td>
<td>frəm, frəm, frəm, frəm, frəm</td>
</tr>
</tbody>
</table>

The above are the possible candidates for the weak form entries for these words.

Since it is impossible to discuss all the other words in the list here, I will give a couple of weakeners which have been found to occur in a very large number of variants. Importantly, in fact, the variations they have are not just for weak forms but also strong, stressed forms.
6.2. “that”

For the word “that,” 351 different forms from a total of 5,871 tokens (excluding three possible mistranscriptions, namely /ɪnɛæt/, /wɪθəɛʔ/ and /UNKNOWN əʔ/) have been found. A summary is given in and below Table 9. Note here that, contrary to Table 8, forms which do occur in Table 1 are in boldface.

Table 9. Strong and weak forms of “that” (n=5,871)

<table>
<thead>
<tr>
<th>Forms occurring 100 times or more (3,812 tokens, about 65% of total)</th>
<th>Number of tokens</th>
<th>Forms occurring 10 times or more (arranged according to their segmental forms; total 1,449 tokens, about 25%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>δæʔ</td>
<td>1,086</td>
<td>Forms occurring 10 times or more (arranged according to their segmental forms; total 1,449 tokens, about 25%)</td>
</tr>
<tr>
<td>δæt</td>
<td>727</td>
<td>δɛʔ</td>
</tr>
<tr>
<td>δɛt</td>
<td>327</td>
<td>δæρ</td>
</tr>
<tr>
<td>δæɡ</td>
<td>203</td>
<td>δɛɡ</td>
</tr>
<tr>
<td>δɛɡ</td>
<td>177</td>
<td>naɛʔ</td>
</tr>
<tr>
<td>naɛt</td>
<td>110</td>
<td>δt</td>
</tr>
<tr>
<td>δt</td>
<td>174</td>
<td>δt?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strong</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>δæ, δɛ, δɛt, δɛʔ, δɪ, δɪd, δɪɡ, δɛ, ɛ, et, eʔ, lɛt, laɛʔ, naɛ, naɛd, naɛr, ɛ, ɛɛ, net, ɛʔ, ɛzɛt, ɛzɛʔ, ɛzɛ, ɛzɛt, ɛθæʔ, ɛθɛ, ɛθɛt, ɛθɛ</td>
<td></td>
</tr>
<tr>
<td>δæ, δɛ, δɛt, δɛʔ, δɪ, δɪd, δɪɡ, δɛ, ɛ, et, eʔ, i, ir, it, iʔ, naɛt, nit, zɛt, zɪt, θɪt</td>
<td>Strong</td>
</tr>
<tr>
<td>δæ, δɛ, δɛt, δɛʔ, δɪ, δɪd, δɪɡ, δɛ, ɛ, et, eʔ, i, ir, it, iʔ, naɛt, nit, zɛt, zɪt, θɪt</td>
<td>Weak</td>
</tr>
</tbody>
</table>

The table comprises roughly 90% of the total occurrences, but there are still 284 forms left which occur fewer than 10 times. I will not divide these into strong and weak forms here:
Note the sheer difference in the number of forms. Another difference is that the most common weak forms have the vowel /i/ here, in contrast to /ə/ in Table 1. This is probably due to the difference in the transcription procedure. According to Flemming & Johnson (2007), reduced vowels in English tend to have [i]-like realization in word-medial positions and [ə]-like realization word-finally. In “that,” the vowel is closed by a consonant (hence word-medial), so [i] might have been interpreted as an instance of /i/ in the process of corpus building.
6.3. “and”

For the word “and,” 189 different forms from a total of 10,998 tokens have been found, which are summarized in and below Table 10. Here again, forms which do occur in Table 1 are in boldface.

Table 10. Strong and weak forms of “and”

<table>
<thead>
<tr>
<th>Forms occurring 100 times or more (9,997 tokens, about 91% of total)</th>
<th>Forms occurring 10 times or more (arranged according to their segmental forms; 723 tokens, 6.6%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forms</strong></td>
<td><strong>Number of tokens</strong></td>
</tr>
<tr>
<td>en</td>
<td>3317</td>
</tr>
<tr>
<td>æn</td>
<td>931</td>
</tr>
<tr>
<td>ēř</td>
<td>475</td>
</tr>
<tr>
<td>ænd</td>
<td>346</td>
</tr>
<tr>
<td>end</td>
<td>294</td>
</tr>
<tr>
<td>ę</td>
<td>173</td>
</tr>
<tr>
<td>æř</td>
<td>159</td>
</tr>
<tr>
<td>en</td>
<td>102</td>
</tr>
<tr>
<td>ų</td>
<td>2165</td>
</tr>
<tr>
<td>in</td>
<td>1085</td>
</tr>
<tr>
<td>on</td>
<td>731</td>
</tr>
<tr>
<td>n</td>
<td>114</td>
</tr>
<tr>
<td>ąř</td>
<td>105</td>
</tr>
</tbody>
</table>

The table comprises more than 97% of the total occurrences, but there are still 151 forms left which occur fewer than 10 times:

α, æαd, æαn, æŋ, ėn, ąnd, ænd, ænds, æņ, æn, æni, ænk, ænm, ænnd, ænndo, ænř, ænz, æn?, æn?, æŋ, ær, æů, æVOCNOISE, ai, ain, ain, aër, an, ąř, au, aon, ąi, ąn, d, źe, den, ą, æænd, ąd, ędn, ąon, ehen, ęi, ęi, ćeŋ, ęŋ, eř, eln, em, ėm, ąm, ąn, ąn, ąnd, end, ęnd, ęnd, ęnde, ęne, ęno, ęno, ęnen, eni, enm, enn, enn, enn, enp, ęnt, enVOCNOISEn, en?
“VOCNOISE” means that vocal noise was present, and “SIL” means “silence”, i.e., the speaker paused in the middle of the word. Compare this again with entries in Table 1.

7. Discussion and conclusion

Obviously, many of the forms found in §6 are the result of some kind of phonological effect other than the lack of accent. Most probably, the initial and/or final consonant change or deletion is influenced by the flanking words to which the word in question is, say, assimilated. In the above, I have not tried to examine the phonetic context of each variant, but I suspect that many of them are the result of such assimilating processes.

Basically, such forms cannot be seen as “weak forms” per se. Instead of listing them all as variants of particular words, I should consider them as the results of a general phonological process of “assimilation.” Other processes which have been identified within phonological theory may also be involved.

However, the forms I have found are so varied that it is possible that current phonology cannot explain them away, and they may reveal processes that have not been identified and incorporated into present phonological and/or phonetic theories. Part of the purpose of this series of studies is to “discover” such processes, but time limitations did not allow me to do that at this time.

I am not sure how far the findings from a given corpus should be incorporated into standard descriptions. The Buckeye Corpus is based on a particular range of speech styles and do not show me the whole picture. But the fact that such a large variation exists is important, and it will be useful to look into this more fully.

It has also been found that not only weak forms but strong forms as well vary a great deal in connected speech. This should be interesting in its own right, both for the description of individual items and the contribution it can have to phonological/phonetic theories.
References


TONIC MISPLACEMENT BY JAPANESE LEARNERS OF ENGLISH

ISAO UEDA AND HIROKO SAITO

Outline

Although not yet discussed widely, one of the distinct characteristics of Japanese-accented English is tonic placement (nuclear stress) on a wrong word in a sentence. The present article reports the result of an experiment to examine what this mis-production is like, in what way the learner's production and phonological knowledge are related, and how the proper accentuation is acquired in the course of time. In the experiment, fifteen university students specializing in English served as subjects. They read and recorded test sentences that included an interrogative word or an attributive adjective and answered which word in each sentence should be pronounced with the highest prominence. To see the development they were asked to repeat the same procedure after one year's interval. The result shows that the subjects can be either correct or incorrect both in production and knowledge. Based on this, we propose a typology of nuclear stress placement consisting of four types. This typology covers each and every case of placement, and furthermore, the acquisition process seems to take place in the form of transition from one type to another. Finally, we conclude the discussion by suggesting that the learner should be formally instructed on tonic placement as well as being practically trained.

1. Overview

It has been widely pointed out that Japanese-accented English exhibits certain notable characteristics including indistinguishableness between /r/ and /l/, merger of (mid-)low vowels, preferred open syllables, etc. The resulting errors of such cases, however, are contextually or pragmatically recoverable most of the time. In this article, we discuss a completely different type of mis-production that can be observed in the highest
prosodic unit, namely the intonation phrase. It is related to the ‘tonic’ placement by Japanese learners of English. The term ‘tonic placement’ is used here to mean the placement of the heaviest stress on a lexical item in a sentence. This phenomenon has seldom, if ever, been a focus of interest and only a few research projects have been conducted. In Japanese learners’ English certain syntactic categories are produced with high pitch and a considerable degree of intensity creating prominence even though the context involves no presupposition or contrast. Because they are ‘prominent’ in the sentence, they may mistakenly be considered the ‘nucleus’ of the intonation phrase. Specifically, the syntactic categories which are typically pronounced this way are personal pronouns, negative particles, interrogatives and attributive adjectives. Let us look at some examples below. (In the following examples and elsewhere, the underlined words indicate the words pronounced with such prominence by Japanese learners of English.)

(1) Four syntactic categories mis-produced with prominence

1 I can play the piano. (Personal pronoun)
2 I haven’t finished my homework. (Negative particle)
3 What would you like to do? (Interrogative)
4 It was in the red book. (Attributive adjective)

Unlike the segmental or syllabic errors mentioned earlier, this type of mispronunciation is not contextually or pragmatically recoverable; far from that, it may be the source of a more serious misunderstanding between the speaker and the hearer. Let us observe this in the following examples.

(2) Example 1

A: Do you play any musical instruments?
B: I can play the piano.

In this dialogue, the way that B answers A’s question, with prominence on ‘I’ might make him sound over-assertive and self righteous, which might in turn cause trouble in the communication between the two people in the long run. By the same token, the statement ‘I haven’t finished my homework.’ (2 in (1)) with nuclear tone placed on ‘haven’t’ typically appears in a discourse such as:

(3) Example 2

A: Let’s go for a swim.
B: I can’t go out till I’ve done my homework.
A: I thought you’d finished your homework.
B: Well, I haven’t finished it yet (so get off my back)!

Thus, uttering statement 2 in (1) without any discoursal presupposition sounds abrupt, or rude, as the case may be.

As was mentioned above, not many studies have been carried out so far, but let us survey what has been argued in these few studies. Watanabe (1994), Date (2003) and Wells (2006) point out certain aspects of this problem and they claim that this is induced by negative transfer from L1 Japanese, where these syntactic categories do receive high-pitched prominence. However, they fail to explain why, among the syntactic categories allowed to be prominent in L1, these four categories alone are mis-produced this way.

Mori (2005) claims that the sentence-initial position is special in that this position renders high pitch to the unit filling it. Moreover, her experiment shows that it is not just any element that can be rendered high pitch in this sentence-initial position. When she made her subjects read out sentences containing mono-syllabic articles (‘a’ or ‘the’) or prepositions in initial position, they were not realized with high pitch by the Japanese learners. Mori concludes that syntactic categories, precisely pronouns (‘I’, ‘they’, ‘it’, and ‘you’), have the greatest effect on the frequency of occurrence of initial high pitch among her Japanese subjects.

Saito (2006) attempts to find the reason why pronouns in sentence-initial position or interrogatives and negative particles in English are incorrectly pronounced with prominence by Japanese learners of English. She suggests that it is negative transfer from Japanese where such words or particles were realized with high-pitch accentuation. She concludes that the accent pattern of L1 was transferred onto the corresponding item of English. However, if this is true, then a speaker with a different L1 accentuation pattern should show a different pattern when they try to speak English. That is to say, speakers of the Osaka dialect, which shows very different accentuation patterns from the standard dialect based on Tokyo speech, do not give high-pitch on some of the interrogatives that the Tokyo dialect speakers pronounce with high pitch, so logically, their English should sound more natural compared to learners of English from Tokyo. To prove this hypothesis, Saito and Ueda (2007) conducted an experiment taking into account dialectal difference in prosody. However, they were not able to obtain statistically significant results, and their claim remains to be a possible suggestion. All in all, no satisfactory explication has been provided up until now.

The most serious problem here seems that these studies are not supported by an examination of a sufficient number of subjects in order to
be considered scientific and systematic, with one exception of Saito and Ueda (2007), in which 38 subjects were examined; they found that this mis-production spreads even across the most advanced learners and that there is a discrepancy between the actual pronunciation and what the learner knows about his/her pronunciation. Although this was the first systematic cross-sectional study, not a single longitudinal study has been conducted thus far.

Given this situation, we attempted a longitudinal study to see how the learners’ phonological system changes in the course of acquisition. Specifically, we examined 15 subjects with respect to their production and phonological knowledge, and after one year’s interval the same experiment was repeated. In this article we report on such an experiment and the result, and discuss related issues, both theoretical and practical. Finally we touch upon some pedagogical suggestions which our argument consequently leads us to. Overall, this article is expected to serve as a preliminary study which will, hopefully, trigger further investigations into this problem.

2. Experiment

The subjects were 15 university students specialising in English at Osaka University of Foreign Studies. At the time of the first data elicitation the students were in their first year, and at the time of collection of the second set of data they had moved on to their second year. During this time, all the students will have studied English from various perspectives in different classes they attended, including practical courses aimed at improving their listening-and-speaking abilities. However, none had spent the year in an English-speaking country, nor were the students given specific instructions on the placement of the nucleus or intonation of English in their phonetics classes. These students are considered to be placed in the topmost rank in terms of English language proficiency in general. The experiment is divided into two parts: a production test and a written test. The exact procedure of the experiment is as follows: first they were asked to read and record the 15 test sentences; after the recording, they were asked to mark one word of each sentence which they thought they should pronounce with the heaviest prominence. In this experiment we focused on two syntactic categories, interrogatives and attributive adjectives. The test sentences are as follows, with the target words underlined and the more ‘correct’ or natural words that should be pronounced with greatest prominence shown in italics. Note that where broad focus is intended, the nucleus usually falls on the primary-stressed
syllable of the last lexical item in a sentence, except when there are words or expressions that denote time, such as today, in which case they will be skipped and a lexical word earlier in the sentence receives the nucleus. There are, of course, cases where there is more than one possible way of reading a sentence without denoting special contrast, and in such cases, the alternative word that might receive stress has also been italicised here. However, none of the interrogatives or attributive adjectives in the sentences presented to the subjects would be given nuclear stress by a native speaker of English unless some kind of emphasis was intended. The sentences were shown to the subjects without underlining or italicisation in the experiment.

(4) Test sentences

1. Where do you live?
2. When were you born?
3. What kind of music do you like?
4. Which class did you attend today?
5. Which book did you borrow?
6. Why did you choose to study at this university?
7. How long have you been studying English?
8. What does your father do?
9. There was an interesting book at the shop.
10. What’s that white bird in the tree?
11. A nice person is more easily fooled.
12. Good wine is, after all, expensive.
13. Which word was in the blue book?
14. Which period are you thinking about?
15. There was a large box in the middle of the room.

The sentences 1-8, 13 and 14 includes interrogatives and 9-15 attributive adjectives as the target. The test sentences include 16 targets in total which are underlined here. As 15 subjects took the test, the total number of the targets amounts to 240. Each of these 240 targets was judged to be either correct or incorrect based upon whether it is properly accented or not.

The recorded production of the subjects was analysed by two trained phoneticians in terms of the location of the nuclear stress in each sentence. Take, as an example, the case of test sentence 1: ‘Where do you live?’. In this case, the targeted item is ‘where’, and if the nucleus is placed on ‘where’, it is judged to be incorrect. On the other hand, if the nucleus is placed on ‘live,’ then it is judged to be correct. If prominence is placed
elsewhere, say, on ‘you’, then it was judged to be irrelevant and it was excluded. Consider another example. In the case of 5: ‘Which book did you borrow?’, if the nucleus is placed on the target ‘which’, then it is judged to be incorrect, while it is judged to be correct if the nucleus is placed on ‘book’ or ‘borrow’. If any of the other words is made the most prominent, then it is judged irrelevant and excluded. Also, in cases where no word in a sentence bears nuclear stress, it is judged irrelevant and excluded from the study.

The phonological knowledge of the subjects was also judged by examining their questionnaire sheets. If the word in a test sentence on which the nuclear stress is placed is correctly marked, it is judged that the subject has correct knowledge of the nuclear stress of the particular sentence. On the contrary, if the subject marked the target word, his/her knowledge is judged to be incorrect. Again, if the subject marks any other word, it is left out of consideration.

3. Results

We might expect the production to be wrong when the knowledge is wrong, and/or vice versa. We expect a kind of ‘symmetry’ between production and knowledge. Contrary to this common expectation, we found an interesting ‘asymmetry’ between them. Production can be either right or wrong irrespective of whether its corresponding knowledge is right or wrong. Also, knowledge can be either right or wrong irrespective of whether its corresponding production is right or wrong. For example,

(5) Subject 11

Production: Where do you live?
Knowledge: Where do you live?

In this case, both production and knowledge are correct. However, in the next case, even though the production may be right, the knowledge is incorrect.

(6) Subject 5

Production: Where do you live?
Knowledge: Where do you live?

As opposed to this, the next example shows that the production is wrong in spite of the correct knowledge.
(7) Subject 2

Production: Where do you live?
Knowledge: Where do you live?

Finally, both production and knowledge may be wrong as in the next case.

(8) Subject 14

Production: Where do you live?
Knowledge: Where do you live?

These four possibilities lead us to propose a typology of the phonology of tonic placement.

(9) Typology of tonic placement

<table>
<thead>
<tr>
<th></th>
<th>Type A</th>
<th>Type B</th>
<th>Type C</th>
<th>Type D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production</td>
<td>Correct</td>
<td>Incorrect</td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Correct</td>
<td>Correct</td>
<td>Incorrect</td>
<td>Incorrect</td>
</tr>
</tbody>
</table>

All the subjects share the above four types in the test sentences. Look at the test result of the sentences 11 through 15 of Subject 12. The underlined words are produced with nuclear stress and the bold words are the ones which the subject answered where the nucleus should be placed.

(10) Test result for Subject 12

11. Production: A nice person is more easily fooled.
   Knowledge: A nice person is more easily fooled.
12. Production: Good wine is, after all, expensive.
   Knowledge: Good wine is, after all, expensive.
13. Production: Which word was in the blue book?
   Knowledge: Which word was in the blue book?
14. Production: Which period are you thinking about?
   Knowledge: Which period are you thinking about?
15. Production: There was a large box in the middle of the room.
   Knowledge: There was a large box in the middle of the room.

In 11, ‘nice’ is mistakenly pronounced with a nucleus, while the subject correctly thinks that it should be placed on the word ‘fooled’. This is Type B. In 12 a wrong word for the nuclear stress ‘good’ is actually produced with the heaviest stress and is also judged that it should be stressed. This is Type D. In 13 ‘book’ is correctly produced with a nucleus in the right place but the subject wrongly answers that the nucleus should
be placed on the word ‘which’. This is Type C. In 14 actual nucleus placement and the answer are both correct, which is Type A. In this way, all our types are observed across all 15 subjects; there is not a single subject that exhibits, for example, Type A alone in all the test sentences. Given this fact, we will only consider the distribution of these types ‘across’ subjects, not ‘within’ subjects.

As has already been stated above, out of a total of 240 targets, we excluded the cases which are irrelevant for the purpose of this study. The total number of targets available is 122. Let us see the number of occurrences of each of the four types, and how they changed during the period of one year.

(11) Number of occurrences of the four types

<table>
<thead>
<tr>
<th>Type</th>
<th>First test</th>
<th>Second test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type A</td>
<td>28</td>
<td>54</td>
</tr>
<tr>
<td>Type B</td>
<td>31</td>
<td>23</td>
</tr>
<tr>
<td>Type C</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>Type D</td>
<td>41</td>
<td>25</td>
</tr>
</tbody>
</table>

If we compare the results of the first and second tests, we may get the impression that the subjects’ tonic placement has been greatly improved. The impression is induced by the fact that, first, the number of Type A subjects has remarkably increased; second, both Type B and Type C students decreased in number, and finally, there is a sharp decrease in the number of subjects belonging to Type D. Note that Type A is perfect, Type B and Type C are partially imperfect, and Type D is the least ideal in respect to tonic placement.

However, when we delve into the change of figures more closely and precisely, we notice that the situation is more complicated than it appears. Let us scrutinize the change type by type.

(12) Type-by-type change

<table>
<thead>
<tr>
<th>First test</th>
<th>Second test</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>B → A</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>C → A</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>D → A</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>D → B</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>D → C</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Unchanged</td>
<td></td>
<td>86</td>
</tr>
</tbody>
</table>

The most notable point here is that out of 122 targets 86 have not changed at all, no matter which type they belonged to at the outset. This is
a strong indication that it is difficult for the learners of English to improve the ability of placing the nucleus correctly within only one year, without formal training. Thirteen targets have changed from B to A, while it is seven C-type targets which changed to A. This study is by no means intended to discuss the change of phonological system from any quantitative viewpoint but does aim at casting light on the relationship between production and knowledge on theoretical grounds. However, so far as the data suggest, it may be that internal incorrect knowledge of nuclear stress is more resistant to change than actual incorrect production. It should also be pointed out that the change of six targets from D to A is limited to only two subjects.

Let us illustrate that two different paths for phonological development are possible based on our proposed typology (9).

(13) Developmental paths

1. Type D → Type B → Type A
2. Type D → Type C → Type A

Saito (2006) mentions one experiment where 22 university students, who showed typical patterns of nuclear-stress misplacement as well as discrepancies between pronunciation and knowledge as to where the intonation nucleus of a sentence should be placed, were subjects of an experiment. These students were given instructions on the rules of English intonation, in particular, on nucleus placement. A month after this short explanation and some practice, the subjects were asked to read the same and similar sentences again, and it was found that there was dramatic improvement in their pronunciation compared to before they were taught the rules. Saito concludes that rules concerning prosodic aspects of language are indeed teachable, and to make learners to become conscious of such rules is pedagogically effective.

To conclude this section, we suggest a pedagogical implication inferred from the discussion. Usually, audio-lingual practice is emphasised in correcting pronunciation. However, the learner sometimes exhibits proper tonic placement but lacks proper knowledge of it: namely, those belonging to Type B of our subjects. Given that such a type exists, and taking into consideration findings from Saito (2006) mentioned above, for some learners formal instruction of tonic placement may be more important and effective than a practice-makes-perfect style of instruction.
4. Summary and discussion

In this paper we have introduced tonic misplacement which is very commonly observed across Japanese learners of English, but has seldom been a subject matter of discussion in second language phonology. This mis-production may result in a misunderstanding in communication at a pragmatic/contextual level. In the light of the fact that no longitudinal research had been carried out previously, we conducted an experiment to observe how proper tonic placement is acquired. We tested how both learners' production and knowledge changed over a period of one year. Based on the result, we proposed a typology in terms of the learner's production and knowledge, both of which may be either correct or incorrect. This led to the learners' tonic (mis)placement being classified into four types. According to the typology, even though the speaker might place a nuclear stress on the proper word in a sentence, the corresponding knowledge may not necessarily be correct. On the other hand, proper placement of nuclear stress superficially does not ensure that the speaker has internalized the correct knowledge. In this sense, learners' phonological systems are not homogeneous with respect to tonic placement. Although the incorrect tonic placement, be it productive or intuitive, tends to become fossilized, there are only two possible paths through which the learner acquires proper nuclear stress placement. We presented a possible pedagogical suggestion based on the discussion.

In the first section of this paper, we pointed out that tonic misplacement may lead to discoursal misunderstanding. Given that misplacement does occur even among advanced learners as we saw from our study, we need, as the next stage, to look into what specific problems that the misplacement may cause in the discourse and to what extent these problems are possibly related to the typology we presented in this study.

Acknowledgments

The present article is based on our presentation at The Second Belgrade International Meeting of English Phoneticians. Our thanks are due to the audience for their comments and questions. We are especially grateful to John Wells and Biljana Ćubrović for their encouraging comments and suggestions. This work was supported by grants-in-aid from Japan Society for the Promotion of Science (20242010 and 20520353 to the first author and 17320083 to the second author), whose support is hereby gratefully acknowledged.
References


PUNCH LINE PARATONE IN ENGLISH
KEN-ICHI KADOOKA

Outline

Paratone is a phonological category which is intended as the counterpart of paragraph in written language. In this paper, I will propose a subtype of Paratone: Punch Line Paratone. Phonologically and phonetically, Punch Line Paratone is realized similar to Low Paratone. I noticed that at the end of jokes – i.e. punch line – the tone of voice is lowered. In my observation, this change of tone is a kind of rhetorical effect to emphasize the termination of a story.

To examine the Punch Line Paratone, an English joke will be analyzed with the intonation notation: 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?” asked the friend. 5) “I send him to my room.”

These five lines consist of six intonation groups, one with a level tone and five falling tones. The falling tones are unmarked while the level tone is marked in this example. It is significant that a brief pause is inserted before the punch line. The beginning of the punch line is indicated by a Low Paratone, which signals the end of this story.

1. Introduction

Paratone is a phonological category which can be regarded as the counterpart of paragraph in written language. One of the most crucial functions of Paratone is that it organizes the structure of the discourse, helping the listener(s) better understand the intention of the speaker. Tench (1996) and Wennerstrom (2001) postulate two kinds of Paratone: high and low. High Paratone marks the change of topics, while Low Paratone suggests the temporary deviation from the main topic.

In this paper, I will propose a subtype of Paratone: Punch Line Paratone. Phonologically and phonetically, Punch Line Paratone is
realized similarly to Low Paratone, which is recognized as a deviation from the main topic in the previous studies. I noticed that it is often the case that at the end of jokes – i.e. the punch line – the tone of voice is lowered to mark the end of the story. In my observation, this change of tone is a kind of rhetorical effect to emphasize the termination of a story. In that sense, Punch Line Paratone may be found in similar kinds of jokes cross-linguistically. Tench (1996: 25) notes that “You should also be able to detect them [phonological paragraphs: KK] in story-telling, and even in jokes.” This paper will be an exemplification of Tench’s suggestion.

Tench (1996: 24) lists the following six characteristics as the phonetic guide to paratone:

(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: KK] between intonation units.

Though the speakers are unconscious about these subtle controls in the intonational patterns, these phonetic modulations undoubtedly signal the structure of stories. These six conditions are considered to match other languages, including pitch-accent languages such as Japanese and tone languages represented by Chinese.

As a framework of the analysis, it will be reasonable to divide the varieties of jokes into two categories cross-linguistically: puns and pragmatic jokes. In this bifurcation, those jokes that are not puns are tentatively defined as ‘pragmatic jokes.’ Puns are language-specific in that they are not translatable; more precisely, when puns are translated into other languages, the meaning of the joke in the original language usually changes. Pragmatic jokes, on the other hand, can be translated into other languages while maintaining the original humour. Below is an example of a pragmatic joke in English (Kadooka 2009). A British speaker read it aloud and I listened to it. The falling tones are indicated with the downwards arrows, and the level tone with a horizontal one. The low paratone is at the beginning of the fifth line. Bold type indicates that the tonic is fallen on those words:
The father was explaining to his friend the difficulty of trying to discipline his son.

“When I was his age, my father sent me to my room for punishment.”

But my son has his own TV, CD player and telephone.”

“So what do you do?” asked the friend.

“I send him to my room.”

These five lines consist of six intonation groups, one with a level tone and five falling tones. Hence, the falling tones are unmarked while the level tone is marked in this example. It is significant that a brief pause is inserted before the punch line.

The humour of this story seems to be universal. Line 5 is the punch line. The difference in the abundance of domestic appliances between the father’s and the son’s room is the crux of the joke. The implication hidden in the punch line is: though the father does not say explicitly, it is implied that there are no TVs, no CD players, no telephones in his room. Hence, it would be a punishment for the son to be sent to his father’s room that has no appliances. It is insightful from the pragmatic viewpoint that the listener of this story immediately understands this implication.

When we testify the six conditions of paratone listed in (1) above to this example, the results are as follows:

1. The high pitch on the onset syllable of the initial intonation unit; the dialogue begins from line 2, the first line being the narration.
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; Father’s comment on the difficulty of the discipline of his son is high in the pitch.
3. There is a gradual lowering of that baseline until the final unit is reached; toward the end of the story, the father’s friend intervenes with the question, which is also uttered in a low tone.
4. The depth of fall in the final unit is the lowest in the whole paragraph; the punch line Paratone is to emphasize the end of the story.
5. There is usually a slowing down process in the final unit. It is also to signal the termination of the story.
6. There is a longer pause than normally allowed between intonation units; the effect of this pause is to attract attention of the listener.

It is certain that all of these six characteristics are realized in the five-line joke. Among these six phonetic characteristics of the Punch Line
Paratone, we will mainly focus on the pause before the beginning of the punch lines hereafter.

2. English jokes

In this section, we will look at how speakers tell jokes. More specifically, we will focus on the intonation patterns of jokes. We will adopt a set of compact discs lasting more than two hours as the source in the following analyses. In these compact discs, several speakers tell jokes, and the laughter by the audience is also recorded.

In a series of the ‘light bulb’ jokes like the one that follows, the structure is very similar; the first line is a question, expecting the negative answer in the second line, and it is followed by the punch line in the third turn. These jokes are the succession of short turns between the two speakers:

(4) ‘Do you know how many Irishmen does it take to change the light bulb?’
   ‘No, I don’t.’
   ‘Fifteen. One’d hold the bulb, and the rest to drink whisky until the room spins.’

(5) ‘How many psychiatrists does it take to change the light bulb?’
   ‘No, I don’t.’
   ‘One, but only if the light bulb really wants to change.’

(6) ‘How many IBM engineers does it take to change the light bulb?’
   ‘I don’t know.’
   ‘Well, none… they really change the standards of darkness and upgrade the customers.’

(7) ‘How many narcissists does it take to change the light bulb?’
   ‘Ah, how many?’
   ‘One, … he holds the bulb and the world revolves around him.’

(8) ‘How many grad students does it take to change the light bulb?’
   ‘I don’t know.’
   ‘One, but it takes ten years.’

(9) ‘How many programmers does it take to change a light bulb?’
   ‘How many?’
   ‘None,… it is a hard work.’
The humour in these stories is based on the ethnic diversity or on the personal peculiarity; the latter can further be divided into the occupation and the mental inclination (narcissists). As the duration time in each story is quite short, the pause posed before the beginning of the punch line may almost be unintelligible. One of the reasons that the pause is short in this series of jokes may be that the pattern or the discourse structure of the three-line turn is so rigid that the punch line in the third line is explicit to the listeners. Another reason, related to the first one, is that these jokes are told by two speakers, hence the short pause before the punch line. It seems that the succession of similar kinds of jokes invokes the laughter of the audience; the faster the tempo, the better; the shorter the intervals between jokes, the more laughter ensues. This is in good contrast with the Japanese *kobanashi* stories to be reviewed in the next section, which are told by one speaker.

Other than the ‘light bulb’ series, there are jokes from various areas. They are not uniformly structured like the ‘light bulb’ ones, but they are shorter with basically two turns. From the viewpoint of the length, the following five jokes are rather ‘minimal’:

(10) ‘Why couldn’t Mozart find his teacher?’
   ‘Because he was Haydn.’

(11) A lawyer sent a note to a client.
   ‘Dear Jim, I thought I saw you on the street the other day, crossed
   over and say hello, but it wasn’t you. So I went back. One tenth of an
   hour, so $25.’

(12) ‘Why won’t sharks eat lawyers?’
   ‘Professional courtesy.’

(13) They say you can pick your nose, and you can pick your friends, but
   you can’t pick your friend’s nose.

(14) ‘Why do men like BMWs?’
   ‘Well.. they can spell it.’

Numbers (12) and (13) are monologues: (12) with a short introduction, and (13) concludes itself. As there is only one speaker from the beginning until the finish of the punch lines of the stories, there is no pause which can be perceived as the paratone. (10), (11), and (14) are dialogues between two speakers. In these three jokes, the pause before the beginning of the punch lines is hardly noticeable.

Maruyama (2002: 214) points out that “In telling jokes, a certain timing and tempo are crucial. Especially, a rapid tempo is necessary for
telling this kind of jokes [about an amnesia patient: KK]” (original in Japanese, my translation). In this sense, the Punch Line Paratone is optional due to the cultural background and the situation where the joke is actually told. This would be a challenge to be pursued further.

It is probable that the intonation patterns are different according to the speaker’s first language. Though told in English, the next examples are by a Japanese speaker (the author herself: Professor Kimie Oshima). The following stories are the direct translation of Japanese kobanashi jokes, hence it is highly probable that the style – e.g. intonation patterns, tempo, tone of voice – is influenced both by the speaker’s first language and the style in Japanese. In the following transcriptions, the pause before the punch lines will be indicated within the square brackets:

(15) Episode 1: Hammer borrower [22 seconds]
Man: Hi, can I borrow your hammer?
Neighbor: No, I don’t want to. I am afraid it will chip.
Man: What? How mean you are! A hammer never chips. It is made of steel. [0.5 second] Well [0.5 second]… fine, I will use mine.

(16) Episode 2: Gold coin [50 seconds]
A man was asleep at night, and he was dreaming. In his dream,
Man: Ohhh, it is really cold today … Hum. What is that? There is an oval gold coin in the small puddle! Wow, I am so lucky! … Ouch! … The water is frozen … I can’t pick it up. What am I going to do? Ah-ha! I know! I can pee on it and melt the ice on the surface. Then I can pick it up!
So the man peed on the puddle and got the gold! Then he woke up.
[0.4 second] The gold was a dream, but the pee was real.

(17) Episode 3: A quack doctor [38 seconds]
Father: Doctor! Doctor! My son fell off the roof! He can’t move his leg!
Doctor: Let me see… Oh, no, it’s broken.
Father: I know. Can you fix it?
Doctor: No, no, it’s already broken. It is too late.
Father: What do you mean … It is too late? He just fell off. I brought him here right away.
Doctor: Yes, but his leg is already broken. It is too late to fix that.
Father: Then when was it not too late?
Doctor: [0.3 second] Well … [0.5 second] before he fell off.

(18) Episode 4: Fishing [53 seconds]
There was an old man fishing at a puddle in front of a bar.
Man: Excuse me, old man, what are you doing?
Old Man: I am fishing.
Man: (Poor old man. He must be crazy.) Would you like to come into the bar with me? I will buy you a beer.
Old man: Oh, thank you.
Man: Two glasses of beer, please! ... Thank you ... Here you go.
Old man: Thank you very much.
Man: So, you said you were fishing. How was it going?
Old man: Oh, it was going very well.
Man: Very well? How many fish have you caught today?
Old man: ha, ha, ha ... [0.4] you are the seventh.

(19) Episode 5: Eel restaurant [60 seconds]
A man lived right next to a grilled eel restaurant. The grilled eels with teriyaki sauce smelt very good while they were cooking. The man could actually eat a bowl of plain rice with the smell. So everyday, he took his rice bowl out in the street and ate his rice in front of the restaurant. One day, the old man who owned the restaurant said,
Old man: Hey, you can’t do that. These are my eels.
Man: Why not? The smell is free for everybody, right?
Old man: OK, you don’t have to pay for the eels, but you have to pay for the smell.
Man: All right. I will bring some money.
He went into his house. Then he brought some change in his hands and shook it.
Man: [0.1 second] There I only smelled your eels, [0.1 second] so you can only hear my money.

Kobanashi stories, as well as Rakugo, are dialogues between two people. In these five examples, four are dialogues and only one is concluded by the narration – the Gold coin story, (16). Including this one, we can observe the Punch Line Paratone in the last lines; there are remarkable pauses of 0.4 or 0.5 seconds, and the lowering of the tone after the pauses. This is in good contrast with the English jokes examined in (4)–(14) without the Punch Line Paratone.

All of the punch lines in these five stories (15)–(19) share the common paratone characteristics defined in Tench (1996). As I listened to the recordings of Oshima (2009), the punch lines underlined in the examples (15) through (19) are realized with the phonetic features of paratone: a slowing down process, a long pause, and the lowering of the tone. Let us
examine the intonation and the paratone patterns with the example (12), which consists of only the dialogue between a man and his neighbor.

(20) 1. The first address by the man to his neighbor begins with high tone.
2. The reply by the neighbor begins with the lower tone than the first address. Then, the second turn of the man is lower than that.
3. The punch line comes in the midst of the man’s utterance. Before and after ‘well,’ there are pauses, the first one longer than the second one.
4. The tone is the lowest in the punch line (in the narrow sense ‘I will use mine.’).

This basic pattern of intonation/paratone is similar to the other five stories in this section. To conclude, the Punch Line Paratone is valid in telling jokes in English.

To conclude, it can be assumed that the insertion or the absence of the Punch Line Paratone, or more specifically the pause before the beginning the punch line, is ascribed to the style of the performance and/or the cultural background.

3. Japanese Kobanashi stories

In this section, the intonation patterns of Japanese jokes will be contrasted with the English counterparts. As two speakers tell jokes in the English examples, it will be meaningful to examine Japanese jokes told by one speaker. The source of the Japanese jokes in this section is the kobanashi – literally ‘small stories’ – which are part of the Rakugo stories.

The ultimate origin of Rakugo goes back to some 400 years ago when story-tellers entertained their samurai feudal lords. During the second half of the seventeenth century in Kyoto and Osaka (Kamigata: literally, upper-region), Rakugo originated in public spaces such as shrines and temples, while in Edo (the former name of Tokyo), it was performed in tatami rooms. In those days, there used to be only shorter anecdotes which lasted a few minutes. About 300 years ago, the format of Rakugo resembled that of the current style. One of the greatest peculiarities of Rakugo is that a single speaker tells the story, performing the roles of all characters.

The first one to be examined is a pun, with the English narration (for the sake of omitting the translation from Japanese) followed by the dialogue in Japanese.
(21) Episode 1: Oil Merchant (*Abura-ya*)

Narration: Once, there was a rumor that a ghost cat haunted on the roof of an oil merchant’s store. The master of the store made up his mind to probe whether the rumor was true or false. At midnight, when he was patrolling, he found the cat and said:

Master: Omae-no-sei-de, uchi no abura wa obaka-abura ya to iwarete, owing to you my store GEN oil TOPIC haunted PARTICLE CITATION say-PASSIVE erai meiwaku shi-teru no ya.

(Owing to you, our oil is dubbed as haunted, and we have much trouble.)

Narration: Then the master threw a stone at the cat. The cat dodged it and said:

Cat: [0.1 Second] Abura-ya no.

The punch line is a pun of “abuna-ya-no” (dangerous) and “abura-ya-no.” In the actual performance, the master’s tone of voice is a little louder than usual, reflecting his excitedness. Then the punch line by the cat is suddenly lowered, falling short of the listener’s expectation which follows the former master’s appeal to the cat. If this punch line were to be uttered similarly to the former excited tone, it would sound out of place.

It is surprising that there are fewer short stories with the pun punch lines than expected, according to the examined material. The next one is one of the few examples:

(22) Episode 2: A Thief and the Deva

Narration: Once, a thief sneaked into a temple, and was going out of the gate with the treasure of the temple. Then the statue of Deva, finding the thief, suddenly began to move and caught the thief. When the Deva trampled the thief, he broke wind and said:

Thief: Niou ka?

Deva Q: ‘Are you the Deva / Do you smell?’

In the actual performance, almost all of this story is the narration, followed by the punch line monologue by the thief ‘Niou ka?’ As well as the former story (7), the punch line follows a longer narration in (8). Then, the next punch line patterns to be examined are the pragmatic ones which are classified as Categories III and IV. The first one is the shortest one, belonging to Category III:
(23) Episode 3: Duck and Scallion (*Kamo ga Negi Shotte*)

Narration: While a duck was swimming in the river, a piece of scallion flew from the upper stream and struck the duck. The duck said:

Duck: [0.1 second] “Oo, kowa...” (How horrible...)

This story is based on the popular Japanese idiom “kamo ga negi shotte yatte-kuru” – literally ‘a duck comes along with a scallion on its back.’ The humour in this story needs a further explanation; in traditional Japanese cuisine, duck is an ideal ingredient, and a scallion is also suitable for the side ingredient, especially when they are cooked together in one pot. Hence, it is a premise that a duck that comes along with a scallion is the most favorable situation for a human being to eat the *kamo-nabe* (duck pot). Under this premise, ducks MUST be afraid of the situation in which they are with a scallion. Hence when a duck happens to come upon a scallion, it would be an ultimate occasion to be afraid of. When a listener calculates this premise and the situation, he/she bursts into laughter.

The next example is a longer ko-banashi with a narration before the monologues of the characters.

(24) Episode 4: Haunted Hole (*Yuurei-no Ana*)

Narration 1: When some youngsters went to a graveyard at midnight, they saw a ghost coming out of a small hole.

1 Youngster 1: Kono ana, husaida-ra donai naru yaro? “If we bury this hole, what will this hole bury-if how become will because of the ghost?”

Narration 2: Then the ghost returns, but can’t find the hole. The day is about to break, so the ghost is upset.

2 Ghost: Washi-no inochi mo koremade ka... “My life seems to be terminated (I am dying...)”

As ghosts are believed to appear only during nights, they will “die” with the daybreak unless they disappear and go back to their netherworld. One of the funny things in this story is that the ghost will die for the second time. When this story is told effectively, line 2 – the punch line – must be said in a pitiful way with a Low Paratone. Though the narration comes before the monologue of a youth and the ghost, the funniest portion is the direct citation of lines 1 and 2.

The second example in this category is longer, with the full dialogue without narration. There are two speakers: Kiroku and Seihachi. In this regard, this is totally different from (11) and (12) with narrations. It can be predicted that the longer the story is, the more necessary rhetorical technique such as Punch Line Paratone is for the listener(s) to understand the structure of the
story, i.e. where is the end of the story. In that sense, Punch Line Paratone may be universal among languages.

(25) Episode 5: Literate Dog

<table>
<thead>
<tr>
<th>Line</th>
<th>Japanese</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Kiroku: Ano yokomachi-wo na,</td>
<td>Listen, around that lane</td>
<td></td>
</tr>
<tr>
<td>2 tooru-no kowai nen.</td>
<td>passing by is terrible.</td>
<td></td>
</tr>
<tr>
<td>3 Seihachi: Dai-no otona-ga,</td>
<td>A grown-up</td>
<td></td>
</tr>
<tr>
<td>4 sonai-na nasakenai koto ii-na.</td>
<td>may not say such a shameful thing.</td>
<td></td>
</tr>
<tr>
<td>5 Kiroku: Soukate, kowai inu-ga otte,</td>
<td>Say, because there is a fearful dog</td>
<td></td>
</tr>
<tr>
<td>6 jiki-ni hoe-kakaru no ya-ga-na.</td>
<td>it always barks at passers-by.</td>
<td></td>
</tr>
<tr>
<td>7 Sehachi: Son-nara, majinai-wo shitaro.</td>
<td>Then, I will give you an incantation.</td>
<td></td>
</tr>
<tr>
<td>8 Tenohira-ni tora-to-iu ji-wo kaiet-yaru sakai,</td>
<td>I will write a Chinese character symbolizing a tiger,</td>
<td></td>
</tr>
<tr>
<td>9 inu-ga hoete-kitara sono ji-wo mise-tare.</td>
<td>when the dog barks at you, let it see the character.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(then the dog will be afraid of a tiger and cease barking; little while after)</td>
<td></td>
</tr>
<tr>
<td>10 Kiroku: Akan, kikime nai de.</td>
<td>No good, it was meaningless.</td>
<td></td>
</tr>
<tr>
<td>11 Yappari hoete-kita sakai,</td>
<td>It barked at me as usual,</td>
<td></td>
</tr>
<tr>
<td>12 tora-no ji-wo miseta kedo na,</td>
<td>I showed it the tiger character,</td>
<td></td>
</tr>
<tr>
<td>13 zen-zen kikime nakatta de.</td>
<td>it was useless, completely.</td>
<td></td>
</tr>
<tr>
<td>14 Seihachi: Uumm, sono inu-wa muhitsu ya na.</td>
<td>Let me see.... the dog must be illiterate.</td>
<td></td>
</tr>
</tbody>
</table>

The whole story consists of only 14 intonation phrases, with a pause inserted in order to suggest that the former half of the lines 1 through 9 and the latter half of lines 10 through 14 take place at different times. The last line, 14, is the punch line. Here the teller lowers the voice and tells more slowly than in the other lines. Such a technique seems to add variation in the course of the
narrative text. The effect of the punch line low paratone is explicit in short stories. If the punch line is told similarly to the other lines, the listeners may have some difficulty to find the punch line.

In conclusion, the tempo in telling jokes in Japanese is much slower than that used in English jokes. Together with the slower tempo, pauses are effective in the art of story-telling, especially before the punch lines. This may be ascribed to the styles or the situation of the performance; in Japanese, it is the tradition to tell Rakugo stories with slower tempo and pauses; in English, on the other hand, the tempo should be faster, with the aim of causing more laughter.

4. Conclusion

In the last two sections, the intonation structures of the English and Japanese jokes were analyzed. As a consequence, it is testified that the following features are common in both languages:

(26) 1. Basically, jokes are told in the form of 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.

There are some exceptions to the first definition that the jokes are dialogues between two people: the monologue of a thief in (5), and that of a duck in (8). The main reason that these jokes are not dialogues is that these jokes are short. In the monologue-type jokes such as (5) and (8) above, the punch lines come immediately after the narrations. Hence, the very beginning of the monologue – it is also the punch line of the stories – starts with the lowest tone. Taking such a specific story structure into account, it must be assumed that the second definition that the dialogue begins with a high tone is abandoned.
References


THE SIGN CHARACTER OF INTONATION

VLADIMIR PHILLIPOV

Outline

Ever since the inception of linguistics as a domain of scientific study, intonation has been its Cinderella, and its status within a multilevel approach to language has invariably been rather tentative. Indeed, some linguists (e.g. Martinet et al.) claim the latter does not belong to langue, yet others (Praguians, Firthians, and generativists inter alia) contend that it is a department of phonology but the contrasts it makes available seem to ‘wander’ across different levels in the research of the different schools of thought.

The present paper views intonation as an exponent of fluctuation, i.e. a shift in the status of a linguistic item leading a different function, while preserving the form. In this respect, it does not merely convey postlexical pragmatic meaning in a linguistically structured way; its substantial characteristics lie in phonology but its formal exponents function explicitly on the morphological, syntactic, and discourse levels. No matter on what level its dominant semantic features operate, it is a Janus-like creature and influences concomitantly all the other levels and sub-levels, too. Its ubiquitous nature makes it a condition sine qua non in the total accountability of the units at any level. Its sign character logically comes from its formal indeterminacy: it is an invariable part of a functioning linguistic sign and it always contributes in a predictable way to its meaning. In this respect it is a co-sign (cf. the Bulgarian при-знак from the German ‘Anzeichen’), or to use Peirce’s term ‘an indexical sign’.

Due to the specificity of its sign character intonation transcends the grammatical and semantic structure of the sentence and steps onto the patterning of the utterance. However, if utterances were to be incorporated in linguistic levels, do we not follow any longer the theory of F. de Saussure concerning his langue–parole dichotomy? Or, rather, do we have to?
1. Introduction

Intonation is commonly equated with prosody or suprasegmentals, the latter including pitch, loudness and speed (or tempo, or speech rate). All these characteristics together form the rhythm of speech and are combined with pauses to chunk the flow of speech. The morphological formation of the term suprasegmental itself lends an erroneous understanding of the concept, i.e. prosody and/or suprasegmentals, and by extension intonation, is viewed as an addition or a superimposed accessory to the spoken language. One is inevitably reminded of the Cheshire cat’s grin syndrome from Lewis Carroll’s classic *Alice in Wonderland*. We all remember Alice’s total bewilderment at the sight of the Cat that vanished quite slowly –

… beginning with the end of the tail, and ending with the grin, which remained some time after the rest of it had gone.

“Well, I’ve often seen a cat without a grin,” thought Alice; “but a grin without a cat! It’s the most curious thing I ever saw in all my life!” (Carroll 1974)

The state of having “a grin without a cat” is certainly the reason why intonation behaves like a ‘floating feature’ – to use metaphorically a concept from Autosegmental Phonology – within the overall description of language. Needless to say, some of the primeval motives go back to the venerable literate tradition of European civilization: Littera scripta manet; ɑɬɨ ɧɚɩɢɫɚɧɨ ɩɟɪɨɦ, ɬɨɝɨ ɧɟ ɜɵɪɭɛɢɲɶ ɬɨɩɨɪɨɦ. Recently, Derrida’s grammatology questioned yet again the overall legitimacy of the subject of phonology and placed writing at the very basis of Western culture. As he claims: “…it could be shown, that there is no purely phonetic writing, and that phonologism is less a consequence of the practice of the alphabet in a given culture than a certain ethical or axiological experience of this practice.” (Derrida 1996: 215).

However, extending Lotman’s theory of the typology of cultures, it can be claimed that speech, and by extension intonation, initially played a more central role in an illiterate culture, “orientated not to the multiplication of the number of texts, but to a constant repetition and recreation of texts… In this case literacy is not necessary. The latter’s role is played by mnemonic symbols … and rituals …” (Lotman 1987: 5).
2. The linguistic tradition in the study of the sign character of intonation

For Saussure, somewhat simplistically, the linguistic signals have available to them only the linearity of time. “In itself, it is merely a line, a continuous ribbon of sound, along which the ear picks out no adequate or clearly marked divisions. In order to do so, recourse must be had to meanings (Saussure 1986 [1916]: 102). The deficiency of the principle related to intonation is all too obvious: the latter is simply left out in the cold.

A lonely voice in the late fifties of the 20th century, which tried to remedy the situation but remained unnoticed in the West, came from the Russian Anglist A. I. Smirnitskij who aptly points that if questions like Has he come? and Is the book on the table? manifest one and the same intonation unit, irrespective of the fact that their sound content is completely dissimilar, then it is better to talk of “bilinearity” (двухлинейность), rather than mere “linearity” (Smirnitzkij 1957: 17). The idea is far from satisfactory but it gets the trend in the right direction: the idea of the mapping of the two lines is already in the air. Unfortunately, Smirnitskij’s untimely death sealed off the fate of that promising intellectual endeavour.

About that time, Charles Bally, one of Saussure’s most eminent students and a co-editor of the Course (together with Albert Sechehaye), was the first to probe explicitly into the sign character of intonation. For him, intonation is the natural expression of modality: “c’est elle qui permet de percevoir si Vous me suivrez est une constatation, une interrogation ou un ordre” (Bally 1950: 42). The situation provides signs which always bear the imprint of reality: they are all actual (actuels). On the contrary –

…une courbe mélodique, un geste, une exclamation, etc., etc. existent sous forme d’empreinte mémorielle, à l’état latent, chez les sujets, et n’entrent en function que dans la parole. Ils ont donc le même caractère que tous les signes proprement linguistiques, d’être virtuels et de devoir être actualisés par l’usage qu’on en fait. (Bally 1950: 44-45)

The idea of the transformation of the intonation curve from a virtual to an actual sign resolves affirmatively the long-standing dispute of whether intonation is a component of silent reading: unless it is, there will be no actualization of its units and the linguistic message will remain deficient. Intonation facilitates the correct reading and comprehension of a written text.
During the late sixties and early seventies the term suprasegmental and its derivatives gained wide currency and the tendency was crowned with scientific legitimacy by Lehiste’s influential Suprasegmentals (Lehiste 1970). Lehiste, an Estonian by birth, had very high reputation both in the East and in the West, and it proved crucial for the ultimate acceptance of the term. According to her, the term suprasegmentals comprises of quantity features, tonal features, and stress features, whose exponents on sentence level are tempo, intonation, and sentence-level stress. To put it differently, these are –

“…features whose arrangement in contrastive patterns in the time dimension is not restricted to single segments defined by their phonetic quality (i.e., distribution of energy in the frequency dimension).” (Lehiste 1970: 3)

It is the lack of spatial and temporal restrictions of the pattern which seems to violate the discrete nature of the linguistic sign and makes any systematic relationship between physical changes and semantic ones well-nigh impossible. Such a line of reasoning makes Labov and Fanshel imply that a search for systematic relationships may be misguided and they suggest that the lack of clarity or discreteness in the intonational signals is not “an unfortunate limitation of this channel, but an essential and important aspect of it” (Labov and Fanshel 1977: 46). The excessive autonomy granted to the suprasegmental level made in its turn its study too haphazard.

The concept of suprasegmental was differently derived in the Eastern European tradition: the oppressively influential Soviet linguistics at the time adopted it as суперсегментный (‘supersegmental’) (cf. Akhmanova 2007 [1966]: 463), pointing to the accusative government of the prefix super, but also to its ablative government, bringing with it the ideas of separation and removal, i.e. intonation can be severed from the segmental level. The term was unanimously accepted and employed by Bulgarian linguists (cf. Tilkov 1981, inter alia).

The turning point in the theoretical development of the subject was the publication of Halliday’s Intonation and Grammar in British English (1967), a work which established the place of intonation within Halliday’s overall approach to language (cf. Halliday 1961). Halliday unequivocally states the principle that “English intonation contrasts [and by extension all ‘stress-accent’ languages’ contrasts – VP] are grammatical; they are exploited in the grammar of the language. The systems expounded by intonation are just as much grammatical as are those, such as tense, number and mood, expounded by other means” (Halliday 1967: 10).
Moreover, “[t]he decision whether a given system that happens to be expounded by intonation is to figure in the grammar or not is grammatical, not a phonological decision; thus it is the requirements of the grammar that set the limits of delicacy on the phonological statement” (Halliday 1967: 11). Halliday also defines the basic discrete unit of intonation, the tone group\(^1\), in terms of rhythmic groups, feet, following in this respect Abercrombie (1964) and thus the tone group is an exponent of grammatical meaning. As T. M. Nikolaeva (1970) remarks in her review of Halliday (1967), the latter follows the Anglo-American tradition of viewing grammar and grammatical meaning as including everything that is meaningful in language. The principle seems to have been initially formulated by K. L. Pike: “In English … an INTONATION MEANING modifies the lexical meaning of a sentence by adding to it the SPEAKER’S ATTITUDE towards the contents of that sentence (or an indication of the attitude with which the speaker expects the hearer to react)” (Pike 1945: 21). Within the continental tendency in intonation research Halliday’s position was rather radical at that time: Soviet linguists were still delineating the subject matter of the discipline and talking of “syntactic phonetics” (cf. Akhmanova et al. 1966), whereas the leading French linguist at the time, A. Martinet, claims that intonation in his understanding, i.e. what is left of the melody once tones and facts of accent are taken from the melodic curve, cannot be analysed linguistically because “the significans does not occupy any given position in the spoken chain but is superimposed, so to speak, on the units of the twofold articulations and we are not in a position to analyse it into a succession of phonemes [i.e. discrete units – VP].” (Martinet 1964 [1960]: 30)

The operational domain of the tone group in Hallidayan grammar is the clause:

There is a tendency for the tone group to correspond in extent with the clause; we may take advantage of this tendency by regarding the selection of one complete tone group as the neutral term in the first of the three systems [i.e. tonality – VP]. That is to say, a clause which consists of one and only one complete tone group will be regarded as “neutral in tonality” (Halliday 1967: 18-19).

It follows logically that –

---

\(^1\) The term was reworded as ‘tone unit’ in later publications (cf. Halliday and Greaves 2008). The latter term is more in congruence with the initial understanding of ‘unit’ in Hallidayan taxonomic grammar (cf. Halliday 1961).
[t]here are then two “marked” possibilities: that the tone group is more than one clause, and that the tone group is less than one clause (Halliday 1967: 20).

Nowadays, the tone group, also called intonation unit (Croft 1995)\(^2\), tone unit (Quirk et al. 1964: 680; Halliday and Greaves 2008: 42), intonation-group (Cruttenden 1986: 35), intonational phrase (Selkirk 1984; Nespor and Vogel 1986), and intonation phrase (Wells 2006), is defined basically in terms of a single intonation contour.

### 3. Diachronic evidence

The question whether there is correlation between intonation units (henceforth IUs) and grammatical units (GUs) has been answered affirmatively in various studies (cf. Quirk et al. 1964 and Croft 1995 *inter alia*), and independent evidence has been adduced and proves that deeper cognitive processes are at work determining the relationship between GUs and IUs. It has also been proved that the upper limit of spoken language is the IU, a feature it shares with the sentence. Thus, it is on the level of the sentence that intonation will display its grammatical, hence sign character at its best. The argument is also in agreement with the traditional view that intonation functions meaningfully on the level of syntax, and within the continental tradition it has always formed one of the integral characteristic features of the sentence, together with predicativity and modality (cf. Durnovo 2001 [1924]: 82-83). In this respect, diachronic evidence can give credence to the argument and can furthermore be a first step towards an analysis of intonation in terms of content typology.

The comparative philology school representatives were – as it is well known – not interested in intonational phenomena, yet some of their sophisticated observations and findings can serve as useful clues as to how the whole complex of prosody functioned in the proto-language. It also serves as independent evidence as to how the process of grammaticalization enhances the functioning of intonation as a grammatical sign.

Following the vein of reasoning of the laryngeal model concerning the Proto-Slavonic vowel system, let us indicate as \( V \) the only proto-vowel, which, consequently, gives rise to the row \( e : o : ė : ď : b \). From the broad root \( *ź \ V \ l – k^w \) -- the verbal-nominal stem \( *ź \ V \ l – k^w – V \) -- is formed by adding the vowel \( V \) (cf. 3 p. sg. present tense \( ěźṭŷ \) from \( *ź \ e \ l – k^w – e \)

\(^2\) Henceforth the term ‘intonation unit’ shall be employed in this study.
– with an e-degree both of the root and of the stem vowel, and nominative, sg. вълкъ ‘an animal which pulls something’ from *ū lb l – kʷ – o – with a reduced degree of the root vowel and an o – degree of the stem vowel). In the Proto-Slavonic sentence *ū V l – kʷ – V - *ū V l – kʷ -- V – ‘the wolf pulls’ the predicate differs from the subject only in terms of the nucleus, that is associated with the former, i.e. the difference is solely prosodic (cf. the nucleus placement in Mod. Bulg. вълкъ влече).

In the proximity of the nucleus the vowel V assumes its highest or e-degree: the latter characterizes exponents of the predicate, i.e. the verbs onto which the nucleus is placed by default (cf. the e-degree in Еλέγετα). Further from the nucleus the vowel V is in its lower (o- or zero) degree; the latter characterizes the exponents of the subject, i.e. the nouns with which the nucleus is never associated with by default. In this way, on the one hand, the verbal-nominal stem *ū V l – kʷ – V – gives rise to the verbal stem *ū e l – kʷ – e - > Еλέγετα carrying a nucleus, and, on the other hand, the nominal stem *ū lb l – kʷ – o - > Елькъ without a nucleus. In the vocative form *ū V l – kʷ – V – the stem vowel V is in the proximity of the nucleus, and since every statement has a predication, the vocative sg. вълкс (= Gk. λύκε, Lat. lupe) the stem V is in its e-degree.3

In conclusion, it can be stated that phonological divergence, and specifically phonemic alterations occur due to the operation of intonation rules within the domain of the sentence. The phonological divergence on its turn brings about the grammaticalization of the focus/ topic bipartition of the sentence. It also lends further support for the phylogenetic evidence of the fact that the correlation between syntactic dependence and prosodic boundary between subject and predicate tends to be a universal.4

4. Synchronic Evidence

Similarly to the processes of word formation, intonation operates across all levels of language (Pencheva 1983), displaying its Janus-like behaviour all the time: the signifier operates on the concrete level but the signified, due to the asymmetric dualism of the linguistic sign (cf.

3 The examples are from Dobrev (1982: 47-48).
4 “Pauses seem typically to occur ... (i) at major constituent boundaries (principally between clauses and between subject and predicate). There is a correlation between the type of constituent boundary and the length of the pause, the more major the boundary, the longer the pause.” (Cruttenden 1986: 37)
Karcevskij 1929), functions on a ‘Chinese box structure’ logic up to the level of discourse, and the whole process can be accounted for in terms of rankshifting of units. With intonation, language in this case performs the phenomenon of fluctuation, i.e. “units from a lower structural level can be integrated into a higher level and vice versa, acquiring features of the latter” (Molhova 1981: 47). Fluctuation can take place only with linguistic signs. Which is the lowest level where fluctuation can occur and intonation operates? It is the word if the ‘Chinese box structure’ principle is to be maintained, but, then, one-word utterances are not interesting unless used for exemplification and teaching purposes. Intonation becomes truly meaningful in the linguistic sense of the word once it becomes the exponent of grammatical relations, irrespective of the fact that it is not a very frequent linguistic phenomenon, occurring as Bolinger aptly put it “around the edge of language” (Bolinger 1964a). Yet, its impact can be quite central: it can express grammatical relations.

In German, intonation can be a grammatical sign for the category of case. In the sentence, Herr Müller schickte dir das Buch und nicht Anna, if the nucleus is associated with Müller, then it is contrasted with Anna, the former being a subject, and the translation 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 the only means of resolving case relation is intonation, since the word Anna being a proper name cannot be (pre)modified by an article, which in German will be an exponent of the case system of nouns. If the word Anna in the latter case 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 type5.

Thus, if it is as Hjelmslev claimed that, “[l]e cas est une catégorie, qui exprime une relation entre deux objets” (Hjelmslev 1935: 96), intonation can be hypothetically viewed as functioning as a marked exponent of case, rather than any of the other grammatical categories. Although at this stage, this is still hypothetical, it has theoretical support from other works that have come out recently in general linguistics (cf. Kasabov 2006).

---

5 The discussion is based on Majtinskaja (1955: 230).
5. Instead of a conclusion

Finally, what is the status of intonation as a linguistic sign? Certainly, it does express emotions, and this aspect falls outside linguistics, but a lot of it does fall, as has been illustrated, within linguistics proper, and linguistics has to provide for the wide variation it can have in its exponence (cf., for example, the many variants of manifestation it can have when expressing emphasis, although the latter conveys only an “immediate emotion” [непосредственное чувство] as Ščerba (1963: 132) puts it).

In 1961 in an ongoing discussion about the nature of the linguistic sign in the philosophical paradigm of the Marxist philosophy of the day, the insightful Bulgarian linguist, the late professor M. Yanakiev defined phenomena which are physically related to the sign as indexical signs, referring explicitly to C. S. Peirce (Yanakiev 1961). If physical relatedness is to be taken as the criteria, then intonation is certainly of this type, for as Bolinger puts it, intonation is “there [in syntax – VP] doing its primitive dance. A lower pitch may be used for theme rather than for rheme; but then themes are generally less interesting than rhemes. It is impossible to separate the linguistically arbitrary from the psychologically expressive” (Bolinger 1964b: 843-844). Thus, intonation seems to be wandering around the category of case as a meaningful indexical signal. In this way it does not lag behind its segmental counterpart but autosegmentally points to it, thus getting rid of the Cheshire cat grin’s syndrome once and for all.

References


PART II:

APPLIED PHONETICS AND BEYOND
THE ROLE OF MODERNIZED PRESCRIPTIVISM IN TEACHING PRONUNCIATION TO UNIVERSITY ENFL STUDENTS

Tvrtko Prčić

Outline

This paper will discuss methods in which modernized prescriptivism could be used in teaching pronunciation to university students of English as the nativized foreign language (ENFL). In the first part, a critical overview of the theoretical and methodological aspects of descriptivism and prescriptivism will be given, followed by an account of the main principles of the newly-proposed modernized prescriptivism. In the second part, methodological and practical pointers on the concrete use of modernized prescriptivism in the teaching of pronunciation will be offered. Finally, the proposal for an integrated approach to teaching language and linguistics, including phonetics and phonology, called ‘usage-enriched descriptivism’, which combines descriptivism with elements of modernized prescriptivism, will be reiterated and reaffirmed.

1. Opening remarks

Building on the ideas in the paper entitled “Descriptivism or prescriptivism in teaching language and linguistics at university level?” (Prčić 2010), this paper is a practical application to the domain of teaching pronunciation of the theoretical and methodological aspects of the concept of modernized prescriptivism put forward there. It is a well-known fact that language and linguistics can be taught either descriptively or prescriptively, and both of these approaches have certain pedagogical advantages and disadvantages. This calls for a critical appraisal of their basic principles and their suitability in teaching language and linguistics in general and, within them, phonetics and phonology, at university level.

The discussion will be organized thus: a summary of the main tenets of descriptivism, prescriptivism and modernized prescriptivism will be
offered in Section 2. Then in Section 3, methodological and practical pointers on the concrete use of modernized prescriptivism in the teaching of pronunciation will be given. And in Section 4, the proposal for an integrated approach to teaching language and linguistics, including phonetics and phonology, called ‘usage-enriched descriptivism’, which combines descriptivism with elements of modernized prescriptivism, will be reiterated and reaffirmed.

2. Descriptivism, prescriptivism and modernized prescriptivism: a summary

In this section, which draws heavily on the formulations in the source paper (Prćić 2010), the basic principles of descriptivism and prescriptivism will be summarized and critically assessed, to be followed by an account of the theoretical and methodological aspects of modernized prescriptivism and an explanation of the needs for its introduction.

According to one view, strongly advocated by the author, to know a language means, firstly, to know the choices it offers for expressing meanings and, secondly, to know how to make appropriate choices from among those available, including the ability to distinguish more acceptable options from less acceptable and unacceptable ones. These kinds of knowledge apply to native and non-native users alike, and especially to university students of language and linguistics, among them to students of ENFL (= English as the nativized foreign language; for overviews of this phenomenon, see Prćić 2003, 2004b, 2005) as future teachers, translators and researchers. In order to provide students with this knowledge, in both of its facets, language and linguistics can be, and indeed mostly have been, taught in conformity with one of two general approaches, largely incompatible and seemingly irreconcilable: one is descriptive in nature and the other is prescriptive.

Descriptivism, or descriptive approach, consists in providing accounts of the language system and, by implication, its use. It is concerned with linguistic resources, i.e. units, on the one hand, and rules and principles for their appropriate combination into higher units, on the other, and can incorporate predictive and explanatory components (for various definitions and interpretations of descriptivism, see Lyons 1968, 1981 and Robins 1989, and the relevant entries in Crystal 2008 and Trask 2007). This approach has three major defining properties: (1) it is objective, because it represents the language as it is, (2) it is synchronic, because it covers, typically, the present time, and (3) it is specific, because it deals with a particular language.
Descriptivism focuses on the possible in language – specifically, on what is grammatical, or well-formed, and is therefore best suited for theoretical courses in all such linguistic subjects which involve inventories, typologies or taxonomies of linguistic resources, at all levels: phonetics, phonology with graphology, inflectional and lexical morphology, syntax, semantics and pragmatics. From the point of view of pronunciation, this approach deals with the phonological subsystem of English and the articulatory, acoustic and auditory aspects of its segmental and suprasegmental components.

However, this is only part of the story, because the use of a language system, and its phonological subsystem is no exception, is not necessarily straightforward and uncontroversial. Rather, it is characterized by variations in form, meaning and style, often lectally and/or registrally determined, and also by misuses by both native and non-native users. These variations and misuses belong to the realm of prescriptivism and usage.

Prescriptivism, or prescriptive approach, consists in providing recommendations on the use of the language system. It is concerned with rules, principles and tendencies of usage, which build around the norms of standard language – and by ‘usage’ it is meant established or customary use (for various definitions and interpretations of prescriptivism, see Lyons 1968, 1981, Milroy and Milroy 1998, Radovanović 2003, 2004 and Robins 1989, and the relevant entries in Crystal 2008 and Trask 2007; and for an excellent review article on prescriptivism and descriptivism in writing on English usage, in Britain, America and elsewhere, from the 18th century to the present, see Peters 2006 and the references therein).

In the past, well into the second half of the 20th century, prescriptivism was authoritarian and subjective, with armchair grammarians pronouncing their dictums, or dicta, on what is correct and what is not, in order to preserve the purity, beauty and literary excellence of the language. This inhibitive ideology, essentially philological in nature, was understandably and rightfully frowned on and revolted against by modern theoretical and general linguists for the most part of the 20th century. Today still, many language teachers pass on their stern judgements on incorrect uses of spelling, pronunciation, grammar and, occasionally, vocabulary, making their students feel discouraged, hesitant and unwilling to use the language for fear of being wrong, regarded as stupid and ultimately ridiculed in class.

Obviously, usage must not be handled in this tactless, black and white manner, nor can it be ignored in the teaching process, as is now frequently the case – incredibly enough, at university level as well, for reasons not
entirely clear at all. Instead, during their education, students of language and linguistics must systematically be familiarized with matters of usage, so as (1) to make them aware of variations of use, (2) to sensitize them to the norms of standard language, both their own and those being acquired as second or foreign languages, and (3) to equip them for making appropriate choices and for recognizing and avoiding inappropriate ones – all this in a deliberate effort to develop students’ attitudes to variations of use and the degree of their acceptability in ENFL usage. And the development of ENFL students’ attitudes to variations in pronunciation should certainly be one of the goals of teaching language and linguistics at university level.

Because of the evident shortcomings of classical, or traditional, prescriptivism, a new, revived, recast, redefined and, above all, destigmatized nature of prescriptivism is needed. The new approach, best called modernized prescriptivism, as it is conceived here, should reflect the achievements of modern theoretical and applied linguistics, and should therefore be no longer philologically, but linguistically, sociolinguistically and corpus-linguistically informed, based not on the notions of correctness, logic, purity, beauty or etymology, but on well-founded and well-explained rules, principles and tendencies of usage, supported by corpus evidence and usage preference polls, all of which would greatly contribute to the objectivity, reliability and authenticity of the recommendations offered.

Modernized prescriptivism should focus on the actual in language, on the realizations of the system – specifically, on what is acceptable on a scale of preferred – less preferred – dispreferred possibilities in accordance with the norms of standard language, when there are cases of variation or misuse. And for pronunciation, standard language means RP in British English and General American in American English (cf. Wells 2008a). Therefore, modernized prescriptivism would be suited for the following aspects of teaching: (1) practical courses in listening, speaking, reading and writing, (2) theoretical and practical courses in translation, (3) theoretical courses in contrastive and contact linguistics, to reduce the effects of negative transfer from L1 at all levels, (4) other theoretical courses, to supplement purely descriptive accounts, and (5) instruction of the norms of standard language in their own right.

3. Modernized prescriptivism and teaching pronunciation

Turning now to the teaching of pronunciation to ENFL university students, there are four areas where recommendations on usage which
comply with the tenets of modernized prescriptivism just sketched out may be particularly appropriate and beneficial:

- the level of phonemes: especially the articulation of individual phonemes,
- the level of words, including proper names: especially the choice of phonemes and stress patterns,
- the level of phrases: especially the stress placement in collocations and idioms, and
- the level of intonation: especially the realization of the fall, rise and fall-rise tunes.

In all these areas, constant attention should be paid to lessening the effects of negative transfer from other languages – specifically, analogies with students’ native language or other languages acquired earlier than English, which would, of course, require a differential or contrastive approach – or, at least, the employment of certain elements of such an approach where and when necessary.

The functioning of modernized prescriptivism in the actual teaching process will now be described and illustrated using the level of words, including proper names, as arguably the most problematic and demanding for ENFL university students. The exemplifications are tailored for students whose native language is Serbian; with minor local adjustments, these examples could easily be adapted for students with different native languages. Once again, the organization, typology and terminology of the system as well as the definitions follow the source paper (Prćić 2010).

As has been originally suggested, recommendations on usage, including those related to pronunciation, can be given in two ways:

- as ADVISORY STATEMENTS, offering guidance on uses which are equally acceptable, with formulations along these lines:
  
  (1) for universally preferred uses: two or more choices are equally acceptable, they are in free variation and may differ only by frequency – use them freely but consistently, e.g. careless as either /'keələs/ or /'keəlis/; suit in British English as either /sjuːt/ or, much less often, /sjuːt/, and

  (2) for restrictedly preferred uses: two or more choices are equally acceptable but they are in complementary distribution in terms of lect and/or register – use them carefully and consistently, depending on your own adopted lect and register, e.g. last in British English as /lɑːst/ and in American English as /laːst/ (the latter is a non-RP variety in
British English); *beret* in British English as /'beret/ or
/bəˈret/ and in American English as /bəˈret/ only;

- as **CAUTIONARY STATEMENTS**, offering guidance on uses which are
  not equally acceptable, some of them being less preferred or
dispreferred, with formulations along these lines:

  1. for less preferred uses: one of the choices is conditionally
     acceptable because... – use it very carefully and sparingly,
     e.g. *mischief* as /ˈmɪʃjɪf/ and *assume* as /əˈʃuːm/ are both
     non-RP varieties, and

  2. for dispreferred uses: one of the choices is unacceptable
     because... – use it at your own discretion, e.g. *asterisk* as
     /ˈæstərɪks/ is a non-standard variety; *faculty* as /ˈfækjʊltɪ/,
     *represent* as /rɪprɪˈzent/ and *Hampshire* as /ˈhæmpʃər/ are
     varieties non-existent in English, essentially false spelling-to-
sound analogies, which are found in the speech of some
foreigners, including those whose native language in Serbian.

Alongside **POSITIVE** formulations of what *can* be done, these
statements may also contain **NEGATIVE** formulations of what *ought
not* to be done, especially to point out systematic, predictable or even expected
misuses among students, some of them stemming from the influence of
their native language, e.g. for speakers of Serbian, *th* is never pronounced
as /t/, except in four proper names: *Anthony* (predominantly in British
English; in American English it is /əθən̩ˈdi/), *Thailand*, *Thames* and *Thomas.*
That is why **ADVANCE RECOMMENDATIONS**, intended for anticipating and
forestalling such misuses, before the actual mistake is made, could
effectively be exploited as well.

With respect to their scope, recommendations can be either **SPECIFIC,**
when taking into account one word only, e.g. the pronunciation of *suggest*
as /ˈsə dʒest/ in British English and as /ˈsə dʒest/ in American English; or
they can be **GENERAL,** when focusing on a range or set of related items,
e.g. the ending *-ford* in names like *Crawford, Hertford, Oxford, Pickford,
Rutherford, Stanford* is never pronounced as /fɔ rəd/ but as unstressed /fəd/;
or, for speakers of Serbian, the medial *s* in words like *basis, hypothesis,*
oasis, philosophy, all borrowings from Classical Greek, is pronounced as
/s/ and not as /z/, as is the case in German and French, through which these
words were borrowed into Serbian in the forms *baza, hipoteza, oaza,*
filozofija.

In class, in the day-to-day teaching process, both advisory and
cautionsary statements can be elicited in at least three ways, all prompted
by what is immediately taught or by what students or teachers have
encountered in films, songs, books, newspapers or elsewhere:
in teachers’ short asides, like these: Note that the word *issue* can be pronounced in three ways – /ˈɪʃuː/, /ˈɪʃjuː/ and /ˈɪʃjuː/; or Mark you that *controversy* can be stressed in two ways – as /ˈkɒntrəvəsɪli/ and as /kənˈtrəvəsɪl/; or By the way, the name of the Scottish capital is pronounced as /ˈɛdɪnbɜːrɡ/ and not as /ˈɛdɪnbaʊɡ/;

in teachers’ or students’ questions asked in class, like these: How do you pronounce the first name of US President Obama – /ˈbɑːrək/ or /bəˈrɑːk/?; or In what ways can the word *garage* be pronounced in British and American English?; or Is the author of *Grapes of Wrath* pronounced (and respelt) as Štajnbek or *Stajnbek* in Serbian?, and

in short thematic research assignments aimed at finding and presenting relevant information, which is provided in print and electronic reference works, on topics like the following: (1) a predictability scale of sound-to-spelling and spelling-to-sound correspondences in English, (2) differences between British and American renderings of phonemes and stress patterns in words, (3) pronunciation of established and recent anglicisms in the students’ native language and of their English models, (4) pronunciation and spelling of countries of the world, continent by continent, and their capitals in English and in the students’ native language, and (5) pronunciation (and (re)spelling) of names of authors, scientists, actors, musicians, etc. in the students’ native language and in English.

Guide to Pronunciation (Olausson and Sangster 2006), BBC Pronouncing Dictionary of British Names (Pointon 1990) and the online howjsay.com: Talking Dictionary of English Pronunciation. And for the treatment of words and names from English in Serbian, the following may come in handy: for anglicisms, Du yu speak anglosrpski? Rečnik novijih anglicizama (Vasić, Prčić and Nejgebauer 2001) and Engleski u srpskom (Prčić 2005); for pronouncing and respelling personal names, Novi transkripcioni rečnik engleskih ličnih imena (Prčić 2008); and for pronouncing and respelling geographical names, Englesko-srpski rečnik geografskih imena (Prčić 2004a).

Of all these reference works, the third edition of the Longman Pronunciation Dictionary (Wells 2008a) offers by far the most useful and usable information for recommendations in accordance with the tenets of modernized prescriptivism. Even though its author has said in an interview (Čubrović 2009: 169) that he sees his job in describing reality as accurately as he can, while leaving teachers to teach EFL, he has nevertheless produced a descriptive masterpiece, in which some of the descriptions can serve as excellent implicit modernized prescriptive (and also remedial or corrective) resources. These features are the following:

- reports, in both numerical and graphical form, on extensive preference polls for pronunciations both within and between British and American usage,
- indications of British and American varieties of words, separated by ‘||’,
- indications of British and American varieties which differ in an unpredictable and important way, marked with ‘(*)’, e.g. geyser as /ˈɡɪzər/ in British English and as /ˈɡaɪzər/ in American English,
- indications of non-RP varieties, marked with ‘§’, e.g. when humid is pronounced as /ˈjuːmɪd/,
- indications of non-standard varieties, marked with ‘!!’, e.g. when nuclear is pronounced as /ˈnjuːkjʊlər/,
- indications of sounds which may be inserted in words, marked with a superscript symbol, e.g. /t/ in prince /prinˈsɪ/ and /s/ in reason /rɪˈzin/., and
- indications of sounds which may be omitted in words, marked with an italicized symbol, e.g. /p/ in attempt /ætˈempt/ and /ɔ/ in mockery /ˈmɒkəri/.

In addition, with a wealth of information made available in the multipurpose electronic Longman Pronunciation Coach (Wells 2008b), of which the Longman Pronunciation Dictionary 3 is a constituent part, together with the Longman Pronunciation Guide, Self Study Lab,
Teachers’ Resource Centre and Pronunciation Preference Polls, and earlier with the Longman Pronunciation Dictionary Study Guide (Fletcher 1990), a companion to the first edition of this dictionary, university students and their teachers are equipped with a complete, reliable, highly motivating and rewarding resource for both descriptive and modernized prescriptive approaches to teaching, studying and self-studying the pronunciation of English.

4. Closing remarks

It must be emphasized, in conclusion, that the role of modernized prescriptivism in teaching pronunciation, in the way conceived and presented here, is definitely not to compete with, let alone displace, descriptivism. Modernized prescriptivism is rather to be seen as a more or less regular complement to descriptivism, where the latter deals with the theory and practice of the phonological subsystem and the former helps in resolving matters of usage of that subsystem, as they arise. Despite the fact that modern linguistics is a descriptive science, teaching language and linguistics, and pronunciation as their module, only descriptively is by no means sufficient, because it disregards important matters of usage and paints only part of the full picture. Likewise, teaching language and linguistics, and pronunciation as their module, only prescriptively, even if it is modernized prescriptivism, is not sufficient either, as it misses the language system and thus also paints an incomplete picture.

As a result, teaching pronunciation – as well as language and linguistics in general – to ENFL university students should be a compromise between two, largely incompatible and seemingly irreconcilable, approaches, provisionally called usage-enriched descriptivism. According to this integrated approach, descriptive accounts of the language system would be combined, where and when necessary, with recommendations on the appropriate use of that system. In the long run, this should result in a better, deeper and more complete acquisition and mastery of language and linguistics in general and of phonetics and phonology as their modules. If, and when, teachers do decide to start enriching – informally or formally, systematically or unsystematically – their classes with apposite tips and pointers on usage, and later to involve their students in short research projects, which the students themselves may initiate if motivated well enough, the full picture of the phonological subsystem of English and its appropriate use, currently quite sketchily painted in the teaching process, will become a more than welcome fact (cf. Prčić 2010).
References


HEARING THE DIFFERENCE:
AN INNOVATIVE APPROACH
TO PRONUNCIATION TEACHING

INGRID PFANDL-BUCHEGGER, MILENA INSAM
AND ISABEL LANDSIEDLER

Outline

This paper presents data from the project FauvoT, an inter-departmental project in the field of foreign language teaching at the University of Graz, Austria (in conjunction with the Technical University Graz and a neuropathic hospital) that focuses on the influence of concentrated listening as a means of accelerating language learning. The aim of the project is to evaluate the efficiency of this innovative approach to language teaching and to consider its suitability for a university context.

The paper discusses the findings and pedagogical implications of students' use of pre-filtered language material, both in the language laboratory and in a home setting. The analysis of the student performance and evaluation suggests that the use of electronically modified tapes provides students with a more focussed perception of foreign language sounds and proves an efficient tool for an effortless acquisition of phonological skills in the target language, in particular for adult learners. As our results show, the technique leads to a clear improvement in students' pronunciation, thus emphasizing the importance of concentrated listening for language learners.

1. Introduction

"Infants learn language(s) with apparent ease [...]" (Kuhl 2009: 837). For adult learners, language learning requires much greater effort, as both learners and teachers will know. Current research in the field of developmental cognitive neuroscience has brought new and significant insights to our understanding of language development. Today's advanced
means of research have shown that children move from a language-universal speech perception towards a language-specific perception determined by their mother tongue. Thus, children become more and more focussed on the sound system of their native language as they grow. This process is connected to a certain neglect of the characteristics of languages other than their mother tongue, which accounts for the increasing difficulty encountered by older children and adults to perceive and produce foreign language sounds.

As early as the 1950s, the French ENT specialist Alfred Tomatis brought forward some advanced ideas on phonological development which closely resemble the current linguistic view. Tomatis held a rather innovative view of language learning by focussing on the restrictions that keep learners from correctly perceiving foreign language sounds and that are due to solidified listening habits formed by our mother tongue. His idea of 'selective hearing' indeed represents the dominant view in cognitive linguistics today and is well worth reconsidering in the light of recent findings.

2. The tenets of Tomatis' audio-phonology theory

Alfred Tomatis was a French ENT doctor who specialised in hearing deficiencies. In his work with singers, he discovered that there was a close connection between hearing and speaking – the frequencies that according to their audiograms his patients could not hear were also missing in their voice spectra. This led him to assume that "the voice can only reproduce what the ear can hear" (1991: 44), in other words, that speech production is closely linked to speech perception. Based on this intrinsic connection between hearing and speaking, Tomatis came to term his theory 'audio-phonology'. With regard to language learning, the theory would imply that pronunciation problems can be due to impaired or faulty sound perception, which in turn can lead to problems with speech production.

Tomatis also found that from early childhood, the ear becomes increasingly sensitive to the sounds of the language to which it is primarily exposed. By thus adjusting its sensitivity to the ambient language, our hearing becomes conditioned to the intonation, rhythm, and predominant frequency patterns of our mother tongue. We develop certain listening habits which, as we grow older, begin to act as acoustic filters that direct our attention to the recognition of sounds that are perceived as relevant for communication (as opposed to mere 'noise'). Such acquired perceptional patterns impair the correct auditory perception of a foreign language.
This auditive selectivity, or selective 'deafness' to foreign sounds as Tomatis called it, is further enhanced by the different distribution of frequency patterns across languages: according to Tomatis, each language favours a specific range of frequencies (based on the specific pattern of harmonics which determine the special quality of a sound) to which the native ear is particularly sensitised. By this, Tomatis does not mean that "there is a deafness to frequencies outside the basic frequency band. But an undeniable sensitivity to certain frequencies exists [...]" (Tomatis 1977/1991:72).

Tomatis then calculated such specific areas of sensitivity by means of envelope curves and called them 'ethnograms'. The following diagram reproduced from Tomatis 1991 displays three of these ethnograms for German, English, and Russian.


As Tomatis' findings are mainly derived from clinical work and as he did not leave any data relevant to his calculations, we tried to find a way of verifying his premise of the ethnograms. In order to demonstrate differences in the frequency ranges across languages and language groups, we compared the spectrograms of a sentence articulated by three female speakers of different language backgrounds.

The three spectrograms below clearly differ in terms of pitch and sound structure, but also in terms of the mean formant frequencies used in the production of identical sounds. 'Formants' constitute the "resonant frequencies of the vocal tract", so-called peaks of acoustic energy in sound waves around a certain frequency that become visible as dark bands in spectrograms (Johnson 2003: 96). In the example above, formant 4 in particular varies among the three speakers and is especially high in the Scottish spectrogram (4560 Hz), of medium height in the American one (4060 Hz), and lowest in the Australian one (3134 Hz). These
variations in formant frequency among Scottish, American, and Australian English support the theory of different frequency range patterns across languages.1

**Diagram 2:** Spectrograms of a sentence uttered by female speakers of Scottish, American, and Australian English.

For adult learners with well-established listening habits, it is often difficult to correctly hear sounds outside their accustomed range of frequencies. When studying a foreign language, they tend to substitute the sounds that are closest to the target in the phoneme inventory of their mother tongue (cf. Kaunzner 1997).

According to Tomatis, selective hearing is, however, a reversible process. After years of research, he found a way to re-train the perceptual

---

1 Tomatis' theory of different frequency ranges across languages has recently been examined in a diploma thesis at the Acoustics Department of the Technical University Graz. The outcomes of the study suggest that it is indeed possible to trace differences in the spectra of French, British English, and German. Due to methodological discrepancies in measuring, a direct comparison between the calculated graphs and Tomatis' ethnograms could, however, not be drawn (cf. Sereinig 2010).
capacity of the ear and improve learners' auditory differentiation by means of a specific listening training that uses electronic modification of sound transmission. For this, Tomatis constructed a so-called Electronic Ear, a machine which emphasises the characteristic traits of the speech sounds of a foreign language and helps to fine-tune our hearing by adapting the sounds to the frequency spectrum of the target language. This fine-tuning is achieved through filtering certain frequency ranges (including an additional gating effect – the transmission of sound is effected through two separate channels with differing degrees of filtering). In addition to improving learners' auditory perception, this 'speech perception training' is also believed to have positive effects on the general comprehension of the target language as well as on oral production skills.

3. Linguistic background

Some of the essential ideas underlying Tomatis' theory of the way we perceive language have been closely investigated in linguistic research over the past few decades. Relevant results with regard to Tomatis' ideas are to be found in the fields of developmental cognitive neuroscience as well as acoustic phonetics and will be discussed in the following.

3.1. Acquisition of phonological competence

The American linguist and researcher Patricia Kuhl, who specialises in the language development of children, has extensively examined the transition children undergo from a universal to a language-specific perception of speech. In her most recent work, Kuhl (2009: 837) writes:

Infants begin life with the capacity to detect phonetic distinctions across all languages, and they develop a language-specific phonetic capacity and acquire early words before the end of the first year.

Kuhl refers to the universal language capacity of children as Phase 1 of their cognitive development. She observes that "this universal capacity is dramatically altered by language experience", and this alteration, according to her investigations, takes place in the second half of the infant's first year of life (Kuhl 2009: 839-840):

By the end of the first year, the infant brain is no longer universally prepared for all languages, but primed to acquire the specific one(s) to which they have been exposed.

She refers to this state as Phase 2 in infant phonetic development.
After children pass this second phase, they perceive foreign language sounds through the listening preferences determined by their native language. The language input that they receive then starts to build up an auditory filter that is focused on the speech sounds and phonemes of their native language. For older children, therefore, and eventually for adults, it becomes difficult to correctly identify slight variations among phonemes that do not occur in their native language. Such unidentifiable phonemes tend to be substituted by similar phonemes existing within their mother tongue. This 'perceptual magnet effect', as Kuhl named this phenomenon (1991), and the underlying ideas are clearly reminiscent of Tomatis' theory of selective deafness to foreign sounds and thus support his view on language acquisition and language learning from the point of view of cognitive neuroscience.

### 3.2. Connection between speech perception and speech production

Another crucial overlap between classical linguistics and Tomatis' ideas comes from the field of acoustic phonetics. The research of the past few decades has identified the important role of speech perception in the context of speech production. This link between perception and production of speech forms the central aspect on which the Tomatis method is based. Alvin Liberman and his colleagues proposed the so-called 'motor theory of speech perception', which identifies the acoustic percept as the motor gesture underlying the articulation of a speech sound. According to the motor theory, we perceive a phoneme by recognising the motor gesture of the vocal tract during the process of articulation. Liberman and his colleagues found that we only perceive speech when certain cortical production sites are activated, and thus argued that the mechanisms of speech perception and speech production are inherently linked, interdependent, and subserved by one neural network. In this sense, they describe the two mechanisms as "only different sides of the same coin" (Liberman & Mattingly 1985: 30).

Patricia Kuhl also discusses how speech perception is important in the context of speech production. She refers to the theory of 'native language neural commitment'. This theory suggests that infants' speech perception abilities indicate, to a certain degree, their later speech production abilities. Kuhl explains that this is due to the fact that "phonetic perception promotes the detection of phonotactic patterns, which advance word segmentation, and, once infants begin to associate words with objects [...], those infants who have better phonetic perception would be expected to
advance faster" (2004: 839). Kuhl and her colleagues found that those infants who could better discriminate phonological units at an early age showed accelerated overall language development in later years. The results of these studies demonstrate that phonetic perception plays an essential role in language acquisition in general.

Given the importance that is attributed to the role of language perception in the field of linguistics, the examined sound perception training is put in perspective as a means to improve language production in learners. The current state-of-the-art in linguistics supports the results of the study discussed in this paper.

4. Project FauvoT

FauvoT (short for German 'Fremdsprachenunterricht mit audio-vokalem Training') is a joint inter-departmental project in the field of foreign language teaching with participants from several language institutions at the University of Graz, Austria, in conjunction with the Technical University Graz and a neuropathic clinic (Landesnervenklinik Sigmund Freud). The project aims at evaluating the efficiency of an innovative approach to language teaching that is based on the audio-phonology theory of Alfred Tomatis. Our intention is to adapt the method to a university context as well as to provide empirical evidence for the influence of focussed listening as a means of accelerating language learning.

In this paper, we would like to present an evaluation of the effects of using pre-filtered language material in the language laboratory with students of Russian and English.

4.1. Method and population

The study was carried out with two different groups of Austrian students of English and Russian over a period of two weeks in two phases. The training sessions were held in a standard language laboratory.

4.1.1. The participants

The participants were Austrian students of English and Russian. The students of Russian were beginners (n=18), who took part in a two week phonetics pre-class (90 minutes per day) to the regular language course for beginners. They were divided into two groups (a FauvoT test group and a control group).
The students of English were advanced learners, who took part in a regular pronunciation class for first year students at the Department of English at Graz University. The students were taught in 2 blocks of 5 consecutive evenings (again, they were divided into a FauvoT test group and a control group).

4.1.2. The method

A regular Tomatis training course consists of a hearing test, followed by individually tailored training sessions with the Electronic Ear. These training sessions include a receptive stage (passive listening to music and language tapes) and an active phase, in which the students learn to control their voices (that are processed through the Electronic Ear and fed back over the headset in filtered form, adapted to the frequency spectrum of the target language).

FauvoT uses a modified approach attempting to align the demands of the method with the timetables and constraints of a university curriculum, trying to reduce the costs of both the expensive electronic equipment and the individualised tutoring. We thus use a standard language laboratory and pre-filtered language materials (instead of individual filtering with Electronic Ears), in order to make it accessible to whole groups of students. The listening training can thus be integrated into regular teaching. Classes usually begin (and end) with listening to music by Mozart (5 to 10 minutes) plus 10 minutes of listening to pre-filtered audiobooks, followed by regular pronunciation training (i.e. working with the laboratory equipment, students listening back to recordings of their own voices). During the listening phases, the lights were dimmed so that the students could relax and concentrate on auditory perception; during the active phases, students were encouraged not to look at their texts but to try to understand and repeat the texts without visual aid or interference.

All audio materials used were pre-filtered according to the settings established by Tomatis for English and Russian. The electronically modified language material was applied both in the language laboratory and for intensive listening in a home setting (students were encouraged to listen to the audio materials that could be downloaded from an e-learning platform).

The students' speaking performance was digitally recorded at the beginning and at the end of the course. An acoustic analysis of selected vowels was produced by means of a specific speech software (PRAAT). Additionally, the students were asked to write diaries and fill in a questionnaire on their experience with the training method.
5. Results and findings

5.1. Russian

The aim of our study was to find out whether Austrian students of Russian would approach the formant pattern of Russian native speakers after completing sound perception training sessions. The hypothesis suggested that the group exposed to filtered audio material (group F) would come closer to the ideal formant structure than the control group that worked with non-filtered material (group N). The analysis of the initial student recordings before their first training session revealed significant differences in the vowel space used by Russian native speakers and by the Austrian students of Russian, which are demonstrated in the following diagram. The vowel positions are depicted in correlation to their articulatory setting, their place of articulation in the oral cavity.

Diagram 3: Vowel positions of Russian native speakers and Austrian students of Russian before the pronunciation class in their articulation of the Russian words 'mila', 'etaž', 'papa', and 'bufet'.

![Diagram of vowel positions in the oral cavity](image-url)
The diagram above illustrates the mean frequency patterns of the first two formants, F1 and F2, of four selected vowels in Russian as spoken by Russian native speakers (outer square) and a group of Austrian students of Russian (inner square). The formant structures depicted exhibit acoustic differences in the pronunciation of identical vowels in Russian native speakers and the Austrian students. It can be inferred from the diagram that Russian native speakers generally produce vowels at more extreme ends of the vowel diagram, while German speakers tend to produce more central vowels. The students primarily used the vowel space that is determined by their native language, Austrian German, which results in a noticeable accent in their Russian pronunciation. Thus, the different frequency values used by the Austrian speakers result from engraving the German frequency patterns onto the Russian language, which leads to their particular 'Austrian accent' in their pronunciation of Russian.

After two weeks of concentrated listening sessions, the students were able to approximate the Russian vowel positions to a great extent. The following diagram contrasts the vowel space of Russian native speakers (outer square), the FauvoT-group (inner dark square), and the control group (inner light square) after completing the pronunciation class.

**Diagram 4:** Vowel positions of Russian native speakers compared to the FauvoT group (group F) and the control group (group N) after the pronunciation class.
The diagram demonstrates that both groups were able to approach the Russian vowel pattern. A comparison between the first and the second student recordings shows that for both the FauvoT group and the control group, the vowel structure clearly moved away from the initially rather narrow German pattern towards the more extended pattern of Russian native speakers. Comparing the progress of the FauvoT group with that of the control group, it can be seen that the FauvoT group approached the overall structure of the Russian vowel pattern more closely than the controls. The FauvoT group was better able to reproduce the positions of the higher front vowels /i/ and /e/. The control group produced /e/ higher in F1 and lower in F2 compared to the Russian model. The Russian formant frequencies of the back vowel /u/ were also better approximated by the FauvoT group, who lowered F2, while the test group largely retained the formant frequencies of the vowel that are more typical of Austrian German.

5.2. British English

The Austrian students of English showed a similar trend in that they approached the vowel pattern of British English more closely after completing the sound perception training sessions. The diagram below shows a group comparison of the development of the vowel /i/ in two groups of Austrian students of English, one of which worked extensively with filtered audio material (FauvoT group 1) and the other which basically followed the regular class with some listening sessions with filtered material on the side (FauvoT group 2) (see diagram 5).

The diagram clearly shows how close the group that underwent extensive sound perception training came to the formant pattern of the British native speakers. Especially in the higher formants F2 to F4, the students were able to increase the frequency to a great extent and thus approach the high frequency values that are characteristic of these formants in the British English /i/ vowel. Also, FauvoT group 1 was better able to imitate the overall structure of the formants in the target language, especially with regard to the abrupt rise from F1 to F2. FauvoT group 2, by contrast, who was considerably less exposed to filtered audio material, largely retained the rather flat plateau between the first and second formants that is characteristic of Austrian German. From these results, it can be concluded that longer and more intense speech perception training sessions lead to significantly better results in the students' pronunciation.
Diagram 5: Comparison of the pronunciation of the vowel /i/ of FauvoT groups 1 & 2 to that of British English native speakers.

The positive impact of sound perception training in the English pronunciation class was further supported by our most recent results, which will be demonstrated by means of a case study of one female student who worked with filtered audio material (enhanced by the gating effect). The following two diagrams demonstrate the development in the student's production of the high back vowel /u/ in the word 'whom' towards the pattern of British English native speakers (see diagrams 6 a&b).
Diagrams 6 a&b: Development of the vowel sound /u/ in 'whom' in one Austrian student of English.
As shown in the diagram, the student was able to adapt the frequency of each formant to the formant frequencies of the British English /u/ vowel. The general progression from F1 to F4 was smoothed and thus approximated the British English formant trend, which is remarkably straight.

Likewise, in the lower back vowel /æ/ in the word 'animal', the student was able to approach the formant pattern of the target language after completing the listening sessions. The following diagram depicts the approximation of F1 to F4 by the student towards the British English formant structure.

The student clearly assimilated the distinct formants to the target frequency of the /æ/ vowel in British English. The approximation of the frequency of each formant, again, moved the entire formant structure in the direction of the structure of the British English vowel.
Diagrams 7 a&b: Development of the vowel sound /æ/ in 'animal' in one Austrian student of English.
6. Discussion

The data analysis shows that speech perception training helped to approximate the frequencies of the first four formants in the Austrian students' pronunciation of both English and Russian vowels. For both groups of learners, a clear tendency towards the formant pattern of the target language could be detected. In those groups that underwent sound perception training sessions, the single formants moved towards the frequency of the formants of the target language. Also, a close assimilation of the general progression from F1 to F4 to that of the target language could be noticed in the selected vowels. Accordingly, the students were able not only to approximate the particular frequency of each formant, but also to better reproduce the overall formant structure of the vowel. From these results, it can be concluded that concentrated listening sessions with filtered language input can enhance students' ability to comprehend and produce sounds in a foreign language, which becomes visible in the altered formant structure of their vowels.

7. Pedagogical implications

In addition to the encouraging findings of the analysis outlined above, results from the student evaluation of the listening training (diaries and questionnaires) also suggest that concentrated listening by means of electronically modified tapes is perceived as an inspiring way of being exposed to a new language by the learners, as a brief look at the questionnaires indicates. According to the evaluated material, students are more willing to use the target language after the training sessions. As can be seen in the chart below, most of them stated that they feel more confident in using the foreign language. All in all, their motivation and pleasure in learning the new language has increased.

**Table 1**: Results from the questionnaires (selected questions)

<table>
<thead>
<tr>
<th>Confidence when using foreign language</th>
<th>Increase in use of foreign language</th>
<th>Degree of motivation and pleasure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher</td>
<td>Yes, more</td>
<td>Higher</td>
</tr>
<tr>
<td>Less</td>
<td>A little more</td>
<td>Less</td>
</tr>
<tr>
<td>No change</td>
<td>No change</td>
<td>No influence</td>
</tr>
</tbody>
</table>
The results of the evaluation of the diaries and questionnaires show that students in general feel positive towards this new method, mainly because it gives them time to listen without feeling pressured. Listening comprehension usually means that students have to listen to a text and simultaneously fulfill certain tasks (fill-the-gap exercises, multiple choice tasks, answering questions, summarizing etc.). Learners have to actively focus their attention on the given information, which renders the listening process highly stressful. Exercises in which students are required to repeat information immediately upon hearing it may put them under considerable pressure as they have to prepare for speaking while still listening. Teachers should be aware that listening in itself is a very complex process, and the main aim of teaching listening skills should thus be to equip students with just those skills necessary for coming to terms with this process.

We therefore suggest incorporating in the curriculum a concentrated phase of passive listening without additional exercises or task fulfilment. This passive listening phase is supposed to give learners time to 'hear' the foreign language, 'tune into' its sounds, rhythm, and intonation patterns, and experience the difference in the overall timbre between their mother tongue and the target language. As one of our students remarked when working with the Electronic Ear, it feels like being born by the current and floating along with the melody ("Man kommt in eine Art Fluss und 'schwimmt' mit der Melodie mit").

The better the students are able to hear the sounds of the target language, the more willing they will be to use the language. In this context, it should be considered that students learn at an individual pace and also have differing learning style preferences. It would therefore seem advisable that listening should be carried out in autonomous learning modules – students could then determine their own speed and timing as well as the number of repetitions. If the learners experience an improvement in their pronunciation, they will feel less inhibited to speak in the target language, which will have a positive influence on their readiness to communicate and in turn on their communicative competence.

With regard to the teaching of listening comprehension, we further propose to make a clear distinction between 'hearing' and listening and include both tasks in the learning / teaching process. 'Hearing' in the sense of simply absorbing the target language without consciously listening should be given sufficient time, not only in beginners' classes, but also in

\[\text{Footnote: For more results on the questionnaires and diaries see Pfandl-Buchegger 2008.}\]
classes for more advanced students. The importance of auditory input for the learning process at all levels is a well-established but often neglected insight (cf. Rampillon 1989: 69). Accordingly, more time should be given to 'hearing' phases, and teachers should promote 'hearing' the target language by providing appropriate material for the learners.

Active listening comprehension followed by specific tasks can be integrated into the language learning process at a later stage. And even at this later stage, passive and leisurely listening should be given its place. Students should be encouraged to listen at all times, which, considering the young generation's habit of using modern audio-technology devices such as MP3 players, should not meet with much resistance. As audio-material is thus more easily accessible, listening should become more enjoyable and therefore more efficient and motivating for the learners.

References


Pfandl-Buchegger, I. 2008. A new approach to the teaching of pronunciation (and language learning) by means of Sound Perception Training: first results from the project FauvoT. In English Language,


QUALITATIVE METHODS IN PHONETIC RESEARCH – A CONTRADICTIO IN ADJECTO?\(^1\)

TATJANA PAUNOVIĆ

Outline

Most phonetic research is still conducted within traditional experimental paradigms, even when 'pedagogically' grounded or undertaken with pedagogical implications in mind. This is true whether researchers examine students' speech production, speech perception, or pronunciation instruction effects. Rarely do we question the methodological frameworks and paradigms we apply, because phonetic research has long been established as an 'exact', 'experimental' and 'quantitative' domain of inquiry. However, I would like to argue that phonetic research would largely benefit from widening the perspective to include qualitative methods of analysis, particularly when conducted in the context of English language learning and teaching.

1. Introduction

For all of us doing phonetic research in the EFL context practical pedagogical implications are, most of the time, 'lurking' behind our phonetic research questions. Issues of EFL pronunciation instruction, assessment, goals, and achievement are very important to us, since we are all more or less pedagogically 'framed' - especially if dealing with more advanced EFL students, and even more so if they are at the tertiary level of education, studying to become language professionals. Our research participants are very often EFL students, and our research is very often action-oriented, hoping to offer insights that could guide teaching practices

\(^1\) This article is part of the project No 178002 supported by the Serbian Ministry of Education and Science.
– whether we specify that explicitly in our research reports or not.

Still, most of our research qualifies as 'proper' phonetic research, and is conducted within traditional experimental paradigms, whether we examine students' speech production, speech perception, or pronunciation instruction effects, and whether we focus on the segmental or suprasegmental level. Rarely do we question the methodological frameworks we apply, because phonetic research has long been established as 'exact', 'experimental' and 'quantitative' research domain.

But, to quote Cohen, Manion and Morrison (2000:3) in their discussion of research methods in the field of education, the choice of research methods is not simply "a technical exercise", a technical matter: "research is concerned with understanding the world [...]", how we view our world(s), what we take understanding to be, and what we see as the purposes of understanding" (Cohen, Manion & Morrison 2000:3).

Therefore, the introduction of qualitative methods in phonetic research would broaden its scope and help us gain a different understanding of the 'world' of speech. If we are to understand the complex phenomena of EFL phonetics and phonology, we need – to borrow Cohen, Manion and Morrison's image (Cohen, Manion & Morrison 2000:3) – more complex lenses and more different kinds of lenses. Laver (2000) pointed out that phonetics is, 'like molecular biology, an interdisciplinary subject par excellence" (Laver 2000:32) – inseparable from physiology or physics, but at the same time a science about people and their behaviour and interaction. Therefore, we should, as wittily remarked by Harre and Secord, "for scientific purposes, treat people as if they were human beings" (Harré & Secord 1972), in phonetics just like in any other social science.

Second, phonetics is a science about human interaction that happens in social contexts. It follows that phonetic study cannot be conducted independently from either the interactional or the social contexts in which speech is observed. "The natural home of spoken language is social interaction", states Local (2003:115). Spoken interaction is, like other human behaviours, a ‘situated activity’ (Cohen et al. 2000:21): people interpret events, contexts and situations, and act on the bases of those interpretations, and their spoken 'activities' should, no doubt, be of interest to phonetics.

2. Qualitative approach

Qualitative approach aims to provide in-depth insights into the studied phenomena, so qualitative data yield detailed 'thick' descriptions (Geertz
To obtain qualitative data we can use a range of design strategies. For instance, naturalistic (anthropological, ethnographic) studies observe real world situations as they unfold naturally, so there are no pre-determined constraints on the data and the research is open to whatever emerges. 'Emergent design' studies are flexible, open to adapting to changed situations and new insights; while 'purposeful sampling' studies focus on the selected cases, participants or events which are most likely to offer rich and illuminative information, since the aim is not to make empirical generalisations from the sample to a wide population, but to gain deeper insight about a phenomenon (Patton 2004:40). In addition to naturalistic and ethnographic research, longitudinal studies, trend studies, and case studies all yield qualitative data, too (Cohen et al. 2000:21).

Qualitative data gathering procedures can be direct (fieldwork observation, in-depth interviews, focus-group interviews, interactive group interviews, dialogues, narratives, and accounts; Patton 2004:4) or indirect (think-aloud protocols or projective techniques such as sentence completion and scenario-, story- or cartoon-completion, free word association). Direct techniques observe and record events and behaviours as they occur, indirect ones observe the effects of those events and behaviours.

In qualitative approaches, data analysis and interpretation are crucial. Data analysis involves searching for recurring patterns and themes, coding, classifying, and re-organizing the data. Interpretation looks carefully for comparisons, causes, consequences, and relationships between items.

Some of the most important features of qualitative data analysis and interpretation pointed out by Patton (2004:41) include the following:

- inductive analysis and creative synthesis, immersion in the details to discover important patterns, themes and relationships,
- a holistic perspective i.e. viewing the investigated phenomenon as a complex system and being sensitive to complex interdependencies and system dynamics,
- sensitivity to the importance of the context (social, historical, temporal).

Finally, from our point of view, another important property of qualitative inquiry is that it is 'triangulated', i.e. conducted in the triangle connecting the researcher, the participant and the audience (Patton 2004:66). Most importantly, the 'reflexive screens' which include culture, age, gender, class, social status, education, family, political praxis, language, values etc. must all be carefully considered during data analysis and interpretation.
2.1. Mixed methodology

Quantitative and qualitative research methodologies rest on two sets of different underlying assumptions, ontological (realist vs. nominalist) and epistemological (positivist vs. non-positivist), transferred to different assumptions about the nature of social reality and human behaviour (Cohen et al. 2000:6). These differences affect not only the choice of the data gathering techniques, but all the fundamental methodological decisions, from the choice of problems, to the formulation of research questions and the interpretation of the data.

However, it is more and more widely recognized that these different approaches can be regarded not as competitive but as complementary, and that, as pointed out by Perry (2005), all research can be placed on a quantitative–qualitative continuum.

Patton quotes Einstein's remark that "not everything that can be counted counts, and not everything that counts can be counted" and concludes that "research, like diplomacy, is the art of the possible" (Patton 2004:12). Morse (1991) pointed out two decades ago that research methodologies are merely tools at our disposal for facilitating our understanding of phenomena. The same belief echoes in Patton's observation that combining multiple methods and kinds of data increases the objectivity of our insight (Patton 2004:252) and enables us to focus on research questions and strategies, rather than framework and paradigm issues.

As a result, in different fields of study, research focus seems to have shifted away from strictly defined paradigm frameworks, towards mixed methodology projects focusing on more practical questions, and not on 'paradigm' limitations. For instance, Patton suggests that when planning our research design we should focus on the purpose, audience and desired outcomes of our research, and that our methodological choices should depend on the kind of questions we want to answer (Patton 2004:13, 213). Accordingly, we should choose the kind of data that would answer our research questions in the best way – qualitative or quantitative data, or both. The focus is on careful planning of the research design so that the role, the purpose, and the relationship of different kinds of data are precisely defined.

Bryman and Burgess (1994:222) point out that qualitative and quantitative data can be related in the same research study in different ways. On the one hand, researchers can use one set of data to, in a way, verify or 'triangulate' the other, e.g. use quantitative data to check the ethnographic data (Bryman & Burgess 1994:223). This is what Morse (1991) refers to as 'between-method triangulation'. In 'simultaneous
triangulation', qualitative and quantitative methods of data collection are used independently, but the findings are complementary; in 'sequential triangulation' the results of one approach help us decide on the next research phase based on the other method of data collection.

On the other hand, there can be a visible division of labour between the two, when they are used to approach the same topic from different angles. Then the purpose is not to 'triangulate' the two sets of data but to allow the quantitative component to map out general patterns and the qualitative phase to reveal details about processes and perspectives of those involved. Or vice versa – depending on our research design.

There are some researchers in the field of education and language learning who endorse the mixed methodology. For instance, what Mason (1994) sees as a major challenge in complex, interdisciplinary and team research studies is precisely linking data from different sources, i.e. quantitative and qualitative data, that is "developing the necessary technical competences to deal with data that have different logical principles". It is important to learn how to move between one set of assumptions and another, since "most of us feel more comfortable in placing ourselves on one side or other of the so-called qualitative-quantitative divide" (Mason 1994:107).

Like Patton and Cohen, Mason concludes that the most important thing in integrating different kinds of data "involves developing mechanisms to ensure that you are asking sensible, meaningful and appropriately limited questions of your data sets" (Mason 1994:109).

While some other authors use qualitative research as a first step followed by quantitative investigation (cf. Poggenpoel et al. 2001; Poggenpoel 2005), Mason points out that the most common misconception about mixed type methodology research is that one type of data is always expected to 'validate' the other set of data. She points out that the two parts of the study need not be designed to validate each other - they can give data on different things and their relationship is not one of confirmation or contradiction, but of complementing (Mason 1994:109).

In language teaching, however, the trend of using qualitative research in the first decade of the XXI century was not impressive, as presented by Richards (2009). In the review of qualitative studies published in 15 journals related to the field of language teaching, Richards reports that in the period 2000 to 2007 in seven of these journals qualitative research reports occupied up to 10% of the published papers, while in four of the journals one out of five published papers employed qualitative methodology, while in only three journals roughly one quarter of the papers involved qualitative research in some form (Richards 2009:152).
These numbers are indicative of a discrepancy between a proclaimed plea for methodological variety and the actual research practices in the field. And if this applies to the broad field of language teaching research, it is even more true of the narrower field of phonetic and pronunciation research.

3. And what about Phonetics?

3.1. Research methodology

As a first step and an illustration to support this point, the methodological design of three recent phonetic studies will be briefly reviewed here, and compared in terms of their defined research aims, research questions, the participants, the data gathering procedures, and the data analysis procedures.

3.1.1. Research study 1

Sounds Serbian? Acoustic properties of Serbian EFL students' speech (Paunovic 2011).

Research aims – to investigate the acoustic properties of Serbian EFL students' vowels. Specifically, we focused on the duration and formant structure of the English vowel contrasts involving the 'short' vowels /i, ʌ, ʊ, u/ and their 'long' pairs, as well as the /e/-/æ/ distinction.

Research questions – 1) Do Serbian EFL students produce phonetic contrasts (F1, F2) between the relevant vowel categories in English? 2) Do they use the phonetic cue of duration to support relevant vowel contrasts?

Participants – twelve 1st-year students of the English Department at the Faculty of Philosophy, University of Niš (average age 19.3, proficiency level B2+ by the Common European Framework of Reference).

Data gathering – three different tasks were used, of varying degrees of control, constraint, and structure, in order to observe how participants used phonetic cues across different speech styles, situations, and contexts.

Task one focused on vowels used in individual words, in citation form. Thirty-two words were selected and illustrated by flashcards. The participants were asked to use the word shown in the flashcard in the carrier sentence 'I see a ____ there'. Each vowel was illustrated at least twice. All the words but five were monosyllabic, and all the vowels observed were in a stressed syllable, in the focus position of the carrier
sentence. The purpose of this task was to provide clear examples of students' vowels in a maximally controlled context.

Task two consisted in reading aloud *The Story of Arthur the Rat*, frequently used for phonetic studies because it offers diverse examples of all the English speech sounds. The story was taken from Markham & Hazan (2002:16) and only slightly adapted. The participants were instructed to relax and read the text naturally, for instance, imagining that they were reading to an audience. The purpose of this task was to provide examples of students' vowels produced in a strictly structured context, but in stretches longer than one utterance. The tokens selected for measurement were all in the focus position of their tone units, some of them were also TU final.

In Task three, the participants were asked to briefly *retell the story from memory*. The purpose was to provide examples of students' vowels used in the least structured context, as similar as possible to spontaneous speech ('semi-spontaneous' speech, according to Markham & Hazan 2002).

Data analysis – speech was digitally recorded, using a Phillips SBC MD650 microphone, directly into the *Speech Filing System* 4.7/Windows 2008 (© M. Huckvalle, UCL). The measurements included vowel formant frequency values (F1, F2), and vowel duration. Formant frequencies were taken from the steady-state portion of the vowel (Peterson & Barney 1952) in the middle third of the vowel duration. All the formant values were first automatically estimated by the program and then re-checked manually. Vowel duration measures were based on manual segmentation, relying on both the waveform and the wideband spectrogram. The data statistically analysed (SPSS 13/Win) and interpreted in terms of average values.

3.1.2. Research study 2

Discourse intonation - Making it work (Paunović & Savić 2009)

Research aims – to examine how Serbian EFL students use intonation to signal information structure and pragmatic meanings at sentence and discourse levels.

Research questions – 1) Which discourse functions of intonation do EFL students signal in a reading task? 2) What phonetic cues do they use to signal discourse functions?

Participants – fifteen 2nd-year students of the English Department, Faculty of Philosophy in Niš (10 female and 5 male, aged 20-21, overall language proficiency level B2+ CEF, 8-12 years of English in formal educational setting, no phonetic training).
Data gathering – a single reading task was used. The 230-word text comprised 42 to 54 possible tone units (TU).

Data analysis – speech was recorded digitally into the Speech filing System 4.6/Windows (© M. Huckvale, UCL). F0 measurements were obtained through three program procedures (F0 track, F0 estimate and F0 autocorrelation) for each individual participant. The measures included F0 level, F0 range (in Hz and semitones, Nooteboom 1997:645; Nolan 2003 774), F0 movement/ direction (F0 contours). The analysis focused on these intonational cues: pitch movement across the tone unit, pitch level at tone unit boundaries (initial and final) and pitch range (the span between maximum and minimum F0 measurement). Pitch movement was transcribed in the traditional 5-tone system (fall-rise \/, fall \, rise / , rise-fall \, level tone —). With respect to the key (H high, M mid, L low), F0 measurements for specific intonational signals were related to each individual participant’s pitch range, measured separately for two parts of the text (narrative/ dialogue). Pitch level at tone unit boundaries was taken to signal transitional continuity (final, continuing, appealing, Chun 2002). Pause duration (in ms) at tone unit boundaries was measured where relevant for the intonation signal investigated. The data were statistically analysed (SPSS 13/Win) and interpreted in terms of average values.

3.1.3. Research study 3

Telling the tale: EFL students' use of prosodic cues in reading and speaking tasks (Paunovic 2009).

The aim of the study – to investigate further some problems identified in previous research (Paunovic & Savic 2009) concerning the use of phonetic cues to signal pragmatic and discourse information, relying primarily on the intonational functions described by Chun (2002) and Vaissiere (2004). The phonetic cues focused in this research were pitch range, key, F0/pitch level and F0/pitch movement within tone units and at tone unit boundaries, relying on Vaissiere (2004), Chun (2002) and Grice and Bauman (2007).

Research questions – 1) How do Serbian EFL students use phonetic cues to signal discourse structure, boundaries and transitions (e.g. to delimit discourse units such as text paragraphs or conversations, tone unit transitions, TU finality, continuation or initiality)? 2) How do they use phonetic cues to signal information structure (new/given information, focus/parenthesis, contrastive focus)? 3) How do they use prosodic cues for interactive and illocutionary functions ('speech act' functions, 'modal'
functions)? 4) Are there important differences in the use of phonetic cues in reading and 'semi-spontaneous' speech?

**Participants** – six 1st-year students of the English Department at the Faculty of Philosophy, University of Nis (average age 19.3; proficiency level B2 CEFR, just starting their introductory course in *English Phonetics and Phonology*).

**Data gathering** – two different tasks were used. **Task one** consisted in reading aloud the slightly adapted *Story of Arthur the Rat* (Sweet 1895), which provided a tightly structured context for the use of prosodic cues and intonation. **Task two** consisted in retelling the story from memory i.e. producing 'semi spontaneous' speech (Markham and Hazan 2002) as the closest equivalent to spontaneous speech in experimental contexts. Our choice of tasks was guided by our research question No. 4 – to compare the use of the same phonetic cues in two different communicative contexts. For instance, research has shown that a wider pitch range is commonly used in reading aloud and acting than in normal conversation; similarly, reading a dialogue is characterised by a wider pitch range than reading a narrative (Chun 2002: 37).

**Data analysis** – the participants' speech was recorded digitally, using the Phillips SBC MD650 microphone, into the Speech filing System 4.7/Windows 2008 (© M. Huckvalle, UCL). The measures taken included F0 level, F0 range (in Hz and semitones, Nooteboom 1997:645; Nolan 2003 774), F0 movement / direction (F0 contours through three program procedures – F0 track, F0 estimate and F0 autocorrelation), and the distribution and duration of pauses (in ms). The measurements were compared for tasks one and two, for each individual speaker and by group means. The data were statistically analysed (SPSS 13/Win) and interpreted in terms of average values.

3.1.4. Looking with a critical eye

We could say that these studies represent average work in our field – they seem to be inherently consistent, complying in their experimental design with the common requirements for acoustic phonetic investigations, and coping as best they could with the common problems in this type of research – a relatively small number of participants; averaging the values (F0, F1, F2, duration) for different speakers; controlling the tempo of speech for different speakers and for the same speaker on different occasions, and the like.

The question, however, is – could they have been "better"? Not in terms of their empirical, quantitative design, but in terms of gaining broader and
deeper insights into the problems investigated, such as the perspective offered by the qualitative approach. In each of the studies presented above, adding a set of qualitative data would contribute to formulating more complete and more insightful answers to the posed research questions. Or they may even lead to formulating more different research questions that would help us understand the problems investigated in a different way.

In all the three studies, whereas the data analysis procedures were undisputable (standard acoustic analysis), the data gathering procedures were, in fact, limited only to controlled speech contexts. In Study 2 only one type of task (reading aloud) was designed to study prosodic cues, in Study 1, focusing on the segmental level, three tasks representing three spoken contexts were used: citation, reading and retelling. In Study 3, focusing on the suprasegmental level, two types of tasks were used – reading aloud and story retelling. However, all these contexts are heavily constrained, not only the citation form of isolated words, but even story retelling. Not one of these contexts was truly conversational or representative of free or spontaneous speech.

Yet, most phonetic research is based precisely on this type of tasks, for several good reasons. First, they are more economical. Second, texts can be adapted to suit the specific focus and aim of the research at hand, and to contain enough tokens of the investigated phenomenon, which would be very difficult to achieve in a less constrained speech situation. Then, we can choose the types of texts our participants feel comfortable with to minimise the stress effects. Finally, highly structured and constrained contexts provide the control we need in order to feel 'comfortable' in our experimental shoes – the studies can be replicated, the quantitative data compared, etc. And all these reasons are justification enough for choosing such a rigid quantitative research paradigm. Moreover, Wennerstrom (1994) points out that “[t]here is a trade-off between oral reading, which allows the researcher to control the content of the text but does not involve the creative function of language, and free speech, which is spontaneous, but does not necessarily yield the desired contrasts in meaning”. She concludes that in her study “the results between the two tasks [were] mostly compatible” (Wennerstrom 1994:419).

However, both our classroom experience and our life experience tell us that students - as indeed all speakers - speak in different ways on different occasions, in different situations, in different contexts, and their 'results' are often not really 'compatible', but quite different. Therefore, the ultimate aim of our phonetic inquiry should be an attempt to understand how students' performance differs on those different occasions, and, possibly, even why, what causes those differences. This may sound like a rather
broad definition of the scope of phonetic research, but the specific context of teaching-oriented phonetic research, performed with a pedagogical perspective in mind, requires such a broad definition of research aims and purposes. That is why in this kind of phonetic research there is plenty of room for asking more different research questions and for seeking the answers through different, combined approaches, quantitative and qualitative.

3.2. Phonetic research with a qualitative perspective

Some recent research that can be described as at least partly grounded in phonetics shows that this can be done. Although such studies are not numerous, they clearly point in the direction of at least three facets where qualitative perspective would be beneficial for phonetically grounded research: 1) naturalistic data collection, 2) careful consideration of the social context in which speech occurs, and 3) understanding the background of spoken interaction – why people do what they do in speech.

Concerning data gathering, samples of naturally occurring spontaneous speech have become much more available over the past several decades. However, Local (2003:115) points out that "despite the advent of large databases of 'spontaneous speech' and an increasing acknowledgement of the relevance of phonetic detail for speech perception and understanding, the [...] phonetics of talk-in-interaction [...] remains largely unexplored". Despite 'dramatic increases in quantitative information about aspects of speech and speech production in connected speech'... 'we still know surprisingly little [...] about the ways in which ordinary people use the phonetic resources of language in everyday talk to undertake interactional tasks' (Local 2003:115, italics mine).

To capture the dynamics of 'interaction phonetics' or 'talk-in-interaction', Local (2003, 2007) has brought together "phonetic analytic techniques and the interactional sequential-analytic techniques of Conversation Analysis" (Local 2003:115), aiming to document systematically "the ways in which speakers and listeners manipulate fine phonetic detail and phonetic variability in producing and interpreting the moment-to-moment flow of everyday conversation". Moreover, he claims that "phonetic aspects of language should in the first instance be understood as shaped by interactional considerations" (Local 2007:1).

A key feature that this approach shares with qualitative approaches based on careful and unbiased observation of naturally occurring data is that it is 'impressionistic': "the approach demands that the analysis prejudges as little as possible the relevance of particular phonetic details
and particular phonetic parameters." This was put forward over 20 years ago by Kelly and Local (1989) in their discussion of the use of general phonetic techniques in transcribing conversational material. Local has often pointed out that "in order to provide a robust account for the organisation and functioning of phonetic detail in everyday conversation we need to: enrich our understanding of 'context' and 'communicative'... and treat all phonetic resources equally and not give analytic privilege to one kind of phonetic parameter over another" (Local 2007:1).

Examples of research studies that did employ 'careful phonetic observation of naturally occurring data' have mostly focused on the interface of phonetics and pragmatics. For instance, Curl (2005) investigated the phonetic properties of repair resolution in spontaneous conversations in American English, reporting that, despite their lexical similarities, the repairs showed two distinct phonetic patterns that corresponded systematically with sequential and interactional (pragmatic) differences.

Curl, Local and Walker (2006) adopt conversation analysis principles but ground their research in acoustic analysis. The authors point out that utterances represent social actions on the part of the speaker, as real as any other forms of behaviour (Curl, Local & Walker 2006:6). They use an array of technical descriptors "to capture observable details of the talk, its sequential organization and the orientation to that organisation displayed by co-participants"

Similarly, Ogden (2006:1752) combines the methodology of conversation analysis with phonetic analysis, i.e. "integrates sequential, interactional, and phonetic analyses to provide an account of how 'paralinguistic' features create meaning". His results support the view that there is a close relationship between the action conveyed in a conversation turn and its phonetic realization, so that the phonetic shape can turn lexically and syntactically shaped agreement into disagreement. Ogden points out that this "highlights the significance of phonetics in participants’ construction of meaning".

Secondly, another way in which the research studies could benefit from qualitative methodology would be by introducing a greater 'sensitivity' to the social context in which the researched phenomena are realized, that is, a careful observation of the context in which the interaction occurs, either by gathering qualitative data through researcher's observation, interviews with participants or other direct and indirect qualitative data gathering procedures, or by broadening the quantitative investigation to include variously defined contexts for data comparison.

This brings to focus sociophonetic research, i.e. the investigations of socially influenced phonetic variations in speech. In their article titled 'The
social life of phonetics and phonology', Foulkes and Docherty (2006) state
that the main task of sociophonetic investigation is to define which social
categories are indexed in speech and how (Foulkes and Docherty
2006:410). The range of social meanings that individuals are able to index
phonetically is very wide, and sociophonetic research shows that social
categories of gender, age, class, and speech style/register are all much
more complex than has been assumed (Foulkes and Docherty 2006:433).
For instance, differences in speech style cover a wide range of features and
systematic variation and may result form a number of sources, such as the
degree of formality, the nature of the topic, the specific audience, the
physical setting in which speech is taking place and the pragmatic
demands of a particular type of interaction.

For instance, in our research studies, the participants were students.
Physically, the research was conducted in the Faculty computer area, in
small groups (2-4 students). The researcher was a non-native speaker of
English, and the participants' teacher. Although these specific
circumstances were duly noted by the researcher, the more important point
here is that the results would certainly have been different had the
researcher been a native-speaker, or somebody other than the participants'
teacher, due to either accommodation effects or social desirability bias, or
some other factors that could have been taken into account as variables. The
participants' vowel qualities, or their turn-taking prosodic signals would
have been different in a different, less formal setting, in a different
communicative situation, in real interaction, etc. The research studies
presented here would have benefited greatly had the investigation included
a set of different contexts and different data, gathered through qualitative
techniques or by broadening the scope of quantitative experimental
situations. It would be important to observe how, why, and to what extent
these differences are manifested in different communicative contexts.

Thirdly, including the qualitative perspective in phonetic research
could help us understand better the background of spoken interaction – why
people do what they do in speech. Traditionally, in phonetic research,
even when naturally occurring speech is investigated, the phonetic form,
the 'how' of speech, is related to very narrowly specified aspects of the
'meanings' expressed in the given situation. We could, however, also ask
why the speakers used such phonetic cues on those particular occasions,
and qualitative interviews could help us gather invaluable information
about the background of spoken behaviour. People have opinions, beliefs
and attitudes, and they influence the way people speak; issues of identity
and social identification are no less important, not only for what our
participants say, but also for how they say it. Understanding these,
although difficult to subsume under 'phonetic' research, would certainly offer us deeper insights into the process of spoken interaction, which is the aim of phonetic investigation. In-depth interviews with the participants, focus group interviews, narrative accounts, think-aloud protocols – all these qualitative data gathering procedures could provide very useful data in this respect, which would be complementary to the data obtained through the quantitative and experimental parts of the research. Therefore, mixed-method phonetic studies could focus on both the how, the what and the why questions by integrating quantitative and qualitative approaches.

4. Conclusion: 'eclecticism and tolerance'

Along the lines discussed above, Miller (2008:653) states that "[a] unique challenge for linguistic inquiry is in designing and implementing analytical method which combines the scrutiny and individualized care of qualitative studies with the replicability, ability to generalize and potential for high numbers of informants of quantitative research". Interdisciplinary and multidisciplinary research projects using mixed methodology seem to be the new, challenging but promising, path of inquiry in all linguistic areas, phonetics included.

But such efforts are still not very commonly put to practice. This is particularly surprising if we keep in mind that it was in the distant 1977 that Dell Hymes wrote: "We have reached, in effect, a study of language that is inseparable from a study of social life, and in which quantitative differences are inseparable from qualitative effects" (Hymes 1977:169).

Similarly, John Laver in his article 'The nature of phonetics' reminds us that the future of phonetics lies in 'eclecticism and tolerance' (Laver 2000: 36). He quotes professor Ladefoged's witty email postscript, that 'the International Phonetic Association is like the Episcopal church - one can hold almost any theoretical position as long as one gets the symbols right', and concludes that "several versions of general phonetic theory [...] co-exist in comfortable agreement [...] in a profession notable for its tolerance of diversity. [...] There should properly be room in the broad and eclectic church of the IPA for many attitudes towards the study of speech" (Laver 2000: 36). Including, hopefully, the use of qualitative methodology in phonetic research, alongside our comfortably familiar phonetic measurements.
References


In the last decades scholars have shared contrasting opinions as to whether differences exist in the amount of speech errors resulting from the administration of different test types (e.g. sentence reading tasks, phrase reading tasks, imitation, free speech) in phonological fieldwork investigation, some claiming differences would occur, others claiming they would not. Already Labov (1966) observed that in the case of adult first language speech “attention varies systematically with different verbal tasks”. Dickerson (1977) noticed similar tendencies in non-native speech. Flege (1987), however, found no general differences relating to test type, yet he did not exclude the possibility that they could be found in children’s speech rather than in that of adults and in relation to individual speech sounds rather than the sound system in general. This paper examines the above mentioned assumptions on the example of the pronunciation of English by Slovene learners in order to establish which one comes closest to reality.

1. Introduction

Fieldwork investigation, especially phonological fieldwork investigation has faced numerous trials when it comes to eliciting the sound production of respondents, be it first language (L1) research or that of the foreign or second language (L2). Numerous articles involving pronunciation have been written, but only few have considered the problems test types might cause in the interpretation of results. Also, some studies have emerged either supporting the importance of selection of test type or refuting it on
the basis of insufficient evidence for its support in order to justify neglecting this issue in past and present research. Hereby, the researcher faces a two-fold problem: s/he can do fieldwork investigation in the shortest time possible by administering compact tests which will elicit the desired results in order to maintain the attention span of the respondents, to ensure the desired results which would in other test types (e.g. *free speech* if the test in question is a reading test) occur or not occur, depending on the whim of the respondent. Studies which are not directly geared at observing differences in responses relating to different test types will often ignore test types altogether. Many will mention the method of obtaining results in their study in passing, or will resort to vague descriptions such as “the tests were administered to the respondents in such and such way …”, or “…the respondents were administered a questionnaire …”, sometimes providing a sample of the questionnaire, but not questioning the fact that the results might have been different if the same samples had been elicited in a different way than by reading from a list of words or sentences. In most cases the administration of the word list is intentional in order to elicit specific phonological targets difficult to obtain in any other way (e.g. Eckman 1987).

Awareness of the fact that studies which do not account for influence of test type might be considered as lacking in validity due to the artificial nature of obtaining results especially holds for speech. Speech at this point is interesting in the sense that it would from among linguistic levels seem least affected by the question of test type. Namely a grammatical, lexical, pragmatic or any other issue, would be difficult, or even impossible, to elicit through reading sentences and word lists, so such strategies would probably be out of the question. Yet in pronunciation, this seems to be a problem, since such test types (namely written ones) lend themselves as an optimal means of obtaining specific results. This, however, does not mean we should neglect the question whether or not to explore all the possibilities relating to eliciting the most natural and spontaneous results. It is obvious that such results could not be obtained through *reading*. Free or spontaneous speech can only be possible if the interviewees were to speak freely in a comfortable and relaxing environment, and better yet, without the awareness of the fact that they are being recorded. In ideal circumstances this would require hours of time, a relaxed atmosphere, prior acquaintance and trust between the interviewer and interviewees, a hidden microphone, etc. But would this not be stretching the issue too far?

A promising solution is offered by modern technology and internet connections which enable masses of data to be stored in a very small space. Researchers from all over the world and from all fields of study can
contribute their recordings to enormous corpora on the Internet. The diverse materials, however, might answer some general questions about language, but they would not give testimony to issues which require uniform conditions such as geographical position (when studying dialects in the L1, or L1 dialect interference in the L2), years of learning added to geographical position (when doing a longitudinal study).

2. English as a second language (ESL) in the *reading* and *free speech* of Slovene learners

Between the years 2002 and 2004, a vast study was performed testing the pronunciation of English in the whole of Slovenia (PES: Jurančič Petek 2007). 35 primary and secondary schools throughout the country participated in the research. They were chosen by means of the ANN (Artificial Neuron Network) System, a statistical programme, which forms smaller groups within a larger unit on the basis of shared common features and defines individual members of these groups to represent the entire sample. Hence the results obtained on a smaller sample can be considered applicable to the larger sample, in the case of this study to the Slovene population. The number of respondents was 287 each being involved in a 10-15 minute interview. The test consisted of reading word lists and sentences, translations into English, free speech and imitation tests. The amount of the phonetically transcribed data and later the statistical data was enormous and the points considered (one of them was, naturally, test type) quite numerous. In the choice of test items, priority was given to words, which would elicit dialect interference on foreign language learning, since Slovenia has quite a number of dialects and the research itself involved schools from all regions. Testing dialect interference was thus a logical thing to do. Also, the concentration span of the respondents was a maximum of 10 to 15 minutes, which is why it was impossible to shape the tests in such a manner that the same word items would occur in the written form, in free speech or translations, and in the imitation test. There were, however, some exceptions, e.g. word items “clothes”, “father”, “girl”, etc. which occurred in more than one test type. Different types of test were nevertheless accounted for in the research, which will be discussed further on in this paper.

The research accounted for ‘correct’ as well as ‘deviant’ responses in the pronunciation of English by young Slovene learners also in order to define positive and negative transfer. It is ‘correct’ responses that provided insight into the effect test types might have on the respondents, while ‘deviant’ responses shed light on the potential sources of interference on
the pronunciation of English by Slovene learners, namely L1 or even dialect interference, interference of different varieties of English (e.g. British English and General American) and an issue relating also to the influence of test type, namely orthographic interference. Positive transfer is much more difficult to detect (due to overlapping with correct target responses), which is why this study paid more attention to the deviant responses, also for didactic reasons. Different types of error can identify different types of influence on foreign language (FL) pronunciation as the source of or cause for error. Once the source is defined, the error is more easily corrected.

PES (Pronunciation of English in Slovenia) was based on Wieden/Nemser’s Pronunciation of English in Austria (1991). Although the latter study made a thorough analysis and well-structured presentation of the pronunciation of young learners throughout Austria on the basis of the production of English in Austria’s four largest cities, it failed to prove one of its initial goals, namely dialect interference. This was probably due to the research being performed in cities rather than in the rural areas. PES drew on this experience, which is why the research concentrated mainly on the rural areas and, consequently, dialect interference was detected and proven.

The Austrian study was designed to test the pronunciation of English on the basis of perception tests, imitation tests and elicitation tests. The authors of the research had enough resources and time to afford elicitation tests based on free speech tasks (mini dialogues, translations, conversation), providing them with most feasible pronunciation results and enabling them to exclude the more “questionable” reading tasks. PES, however, due to lack of funding, human resources and time could not afford to exclude reading tests from the tasks, even though these could on occasion give pronunciation results affected by orthographic interference, but on the other hand enabled us to test which test type, if any, might give better results.

3. The relevance of ‘correct’ and ‘deviant’ responses

Different studies require eliciting and interpretation of different results. In earlier phases of linguistic investigation ‘correct’ results were observed in order to structure language systems and compare them among each other in order to learn more about them. This was in line with universal grammar, whether it was dealt with from the structuralist or the generativist point of view, particularly in the latter one, since it is here
“assumed that there is only ONE correct grammar … [and it is] … this one grammar that we call “psychologically real””. (Hyman 1970:59)

Later, that is in the second half of the 20th century, when languages other than the fully developed adult speech where only language patterns and rules were taken into consideration, e.g. child language, dialects, language of the speech and hearing impaired, etc. became centre of attention, the focus diverted from ‘correct’ results to ‘deviant’ or ‘erroneous’ responses to identify the deviations appearing in these languages and rectify them. Unlike ‘correct’ responses which existed for a sake of their own and at this point had nothing else to offer, ‘deviant’ or ‘erroneous’ responses provided an entire range of opportunities to explore the different sources of error (through comparison), identify them and eventually propose remedial procedures for their elimination. In the framework of Contrastive Analysis (CA) in second language acquisition (SLA) ‘deviant’ responses were predicted as a result of comparing different language systems, while in the Error Analysis (EA) that followed they emerged as a natural result of applying the more theoretical CA into practice (James 1998).

Linguistic fieldwork investigation was thus geared towards predicting, observing and analyzing ‘erroneous’ responses, especially as SLA studies developed as a basis for studying the development of language and seeking strategies for the improvement of second language learning.

Although ‘correct’ results are in fieldwork research frequently more numerous and usually the basic goal of a study (we do want to know how much our respondents have learned), they are the least interesting for analysis and interpretation. Due to their uniform nature, it is difficult to distinguish even something as significant as positive L1 transfer from regular L2 scores. It is usually from the ‘erroneous’ responses (even though we together with the ‘correct’ results treat them as part of the learner’s interlanguage and they are not considered something negative but rather a developmental stage in SLA) that we learn most about the sources of error, which in effect helps us to develop strategies to avoid such responses and increase the amount of correct responses which is undoubtedly our final goal.

The present paper places ‘correct’ responses rather than the ‘deviant ones’ in the centre of attention, even if only to demonstrate the importance of choosing the appropriate test type to elicit the results of interest. The fact of the matter is that although the test results in the study of the pronunciation of English in Slovenia show that in most cases ‘correct’ results scored higher in free speech than in reading, the point is not where
higher scores are obtained but whether there is any significant pattern at all relating to test types.

Accordingly two hypotheses are presented and dealt with further on in the paper, namely:

Hypothesis 1: Different test types can /do not elicit different results in terms of correctness

Hypothesis 2: If the first hypothesis holds, this would affect the results of studies, especially if they focus on the level of achievement in SLA (namely, by manipulating the tests, better results can be achieved).

4. Reasons for interpretation of the ‘correct’ results in the PES and theoretical justifications

Tables including correct responses across speech sounds, test types and age groups:

Table 1

<table>
<thead>
<tr>
<th></th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
<th>/b/</th>
<th>/d/</th>
<th>/g/</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-R*</td>
<td>17.6</td>
<td>9.3</td>
<td>56.8</td>
<td>56.8</td>
<td>45.2</td>
<td>65.6</td>
</tr>
<tr>
<td>5-FS</td>
<td>-</td>
<td>3.2</td>
<td>59.6</td>
<td>76.1</td>
<td>59.3</td>
<td>58.6</td>
</tr>
<tr>
<td>5-I</td>
<td>4.9</td>
<td>-</td>
<td>41.0</td>
<td>-</td>
<td>17.1</td>
<td>-</td>
</tr>
<tr>
<td>7-R</td>
<td>20.9</td>
<td>9.8</td>
<td>62.6</td>
<td>82.8</td>
<td>34.2</td>
<td>75.0</td>
</tr>
<tr>
<td>7-FS</td>
<td>-</td>
<td>3.6</td>
<td>46.1</td>
<td>90.2</td>
<td>45.3</td>
<td>75.0</td>
</tr>
<tr>
<td>7-I</td>
<td>4.7</td>
<td>-</td>
<td>43.3</td>
<td>-</td>
<td>15.3</td>
<td>-</td>
</tr>
<tr>
<td>I-R</td>
<td>22.1</td>
<td>7.8</td>
<td>59.3</td>
<td>92.9</td>
<td>46.8</td>
<td>70.8</td>
</tr>
<tr>
<td>I-FS</td>
<td>-</td>
<td>5.1</td>
<td>65.6</td>
<td>87.8</td>
<td>57.1</td>
<td>82.1</td>
</tr>
<tr>
<td>I-I</td>
<td>12.5</td>
<td>-</td>
<td>47.5</td>
<td>-</td>
<td>29.2</td>
<td>-</td>
</tr>
<tr>
<td>III-R</td>
<td>22.2</td>
<td>18.1</td>
<td>68.7</td>
<td>85.7</td>
<td>39.6</td>
<td>71.2</td>
</tr>
<tr>
<td>III-FS</td>
<td>-</td>
<td>5.6</td>
<td>55.6</td>
<td>90.5</td>
<td>66.7</td>
<td>88.9</td>
</tr>
<tr>
<td>III-I</td>
<td>4.5</td>
<td>-</td>
<td>56.5</td>
<td>-</td>
<td>43.5</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1: Scores (in %) for correct responses for voiceless plosives in initial position in word items (poor, pet, pest, pair, puts and /pikætɔrɪ/; ten, two) tend, table, television; (cat, cold, captain, cart), cost, combat, cat, /keɪkæz/ and /kæʃlɪ/; and voiced plosives in final and initial position in word items stab, job, rubs, /ɔræbz/; (played, food, good, head, slide, weeds) lid, words, /ruːdʒd/; geese, garden, girl. *R stands for reading, FS for free speech and I for imitation across the 5th and the 7th grade of primary school and the 1st and 3rd year of secondary school.
Table 2

<table>
<thead>
<tr>
<th></th>
<th>/f/</th>
<th>/θ/</th>
<th>/s/</th>
<th>/ʃ/</th>
<th>/h/</th>
<th>/v/</th>
<th>/ð/</th>
<th>/z/</th>
<th>/dʒ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-R</td>
<td>97.7</td>
<td>23.9</td>
<td>95.3</td>
<td>70.1</td>
<td>3.8</td>
<td>34.4</td>
<td>5.3</td>
<td>17.4</td>
<td>51.2</td>
</tr>
<tr>
<td>5-FS</td>
<td>-</td>
<td>22.5</td>
<td>-</td>
<td>84.6</td>
<td>-</td>
<td>44.7</td>
<td>16.3</td>
<td>31.8</td>
<td>31.6</td>
</tr>
<tr>
<td>5-I</td>
<td>-</td>
<td>16.8</td>
<td>36.1</td>
<td>90.2</td>
<td>14.6</td>
<td>-</td>
<td>0.0</td>
<td>14.8</td>
<td>-</td>
</tr>
<tr>
<td>7-R</td>
<td>100.0</td>
<td>24.1</td>
<td>98.8</td>
<td>92.0</td>
<td>7.0</td>
<td>40.1</td>
<td>10.7</td>
<td>27.9</td>
<td>73.6</td>
</tr>
<tr>
<td>7-FS</td>
<td>-</td>
<td>33.7</td>
<td>-</td>
<td>100.0</td>
<td>-</td>
<td>47.1</td>
<td>12.7</td>
<td>31.4</td>
<td>29.4</td>
</tr>
<tr>
<td>7-I</td>
<td>-</td>
<td>25.2</td>
<td>36.1</td>
<td>98.8</td>
<td>13.1</td>
<td>-</td>
<td>0.0</td>
<td>14.1</td>
<td>-</td>
</tr>
<tr>
<td>I-R</td>
<td>100.0</td>
<td>42.9</td>
<td>100.0</td>
<td>97.6</td>
<td>5.8</td>
<td>55.3</td>
<td>13.1</td>
<td>50.0</td>
<td>78.6</td>
</tr>
<tr>
<td>I-FS</td>
<td>-</td>
<td>50.0</td>
<td>-</td>
<td>100.0</td>
<td>-</td>
<td>74.4</td>
<td>15.4</td>
<td>53.6</td>
<td>56.3</td>
</tr>
<tr>
<td>I-I</td>
<td>-</td>
<td>30.4</td>
<td>37.5</td>
<td>95.8</td>
<td>16.7</td>
<td>-</td>
<td>22.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>III-R</td>
<td>100.0</td>
<td>60.7</td>
<td>96.3</td>
<td>98.8</td>
<td>7.1</td>
<td>55.0</td>
<td>15.5</td>
<td>39.3</td>
<td>78.2</td>
</tr>
<tr>
<td>III-FS</td>
<td>-</td>
<td>51.9</td>
<td>-</td>
<td>100.0</td>
<td>-</td>
<td>41.0</td>
<td>35.7</td>
<td>40.7</td>
<td>31.3</td>
</tr>
<tr>
<td>III-I</td>
<td>-</td>
<td>48.9</td>
<td>47.8</td>
<td>100.0</td>
<td>8.7</td>
<td>-</td>
<td>0.0</td>
<td>19.6</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: Scores (in %) for correct responses for voiceless and voiced fricatives and voiced affricate in word items food; thirteen, thank you, /θtʃl/, /θiːtʃ/; stranger, /strɪkə/; share, _ship, sugar, wash, /pɪkʃərɪ; _hens, _head, /hɛmt/; _lɪe, _loʊe, _lɛɪv, _dɪv, _hæv, _æktɪv, _fɪv, _hænz, _hæd, /Jʊv/; liv, _lʌv, _lɛɪv, _dɪv; (those, the, there, this) them, _ðens, /tʃəkə/; trees, _trees, /kɛtʃəz/; badge, _brɪdʒ in different test types and across age groups.

Table 3

<table>
<thead>
<tr>
<th></th>
<th>/u/</th>
<th>/uː/</th>
<th>/iː/</th>
<th>/ɪ/</th>
<th>/e/</th>
<th>/æ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-R</td>
<td>33.3</td>
<td>20.6</td>
<td>51.4</td>
<td>20.9</td>
<td>50.0</td>
<td>36.0</td>
</tr>
<tr>
<td>5-FS</td>
<td>15.8</td>
<td>43.8</td>
<td>62.0</td>
<td>50.0</td>
<td>62.3</td>
<td>64.8</td>
</tr>
<tr>
<td>5-I</td>
<td>-</td>
<td>50.6</td>
<td>48.6</td>
<td>26.5</td>
<td>67.0</td>
<td>33.8</td>
</tr>
<tr>
<td>7-R</td>
<td>50.8</td>
<td>28.6</td>
<td>54.9</td>
<td>29.6</td>
<td>52.2</td>
<td>46.5</td>
</tr>
<tr>
<td>7-FS</td>
<td>38.5</td>
<td>41.5</td>
<td>67.3</td>
<td>60.5</td>
<td>63.7</td>
<td>60.5</td>
</tr>
<tr>
<td>7-I</td>
<td>-</td>
<td>54.4</td>
<td>45.8</td>
<td>28.5</td>
<td>63.4</td>
<td>42.0</td>
</tr>
<tr>
<td>I-R</td>
<td>61.8</td>
<td>27.3</td>
<td>72.2</td>
<td>34.8</td>
<td>45.0</td>
<td>57.1</td>
</tr>
<tr>
<td>I-FS</td>
<td>15.8</td>
<td>55.0</td>
<td>75.0</td>
<td>67.9</td>
<td>57.5</td>
<td>71.4</td>
</tr>
<tr>
<td>I-I</td>
<td>-</td>
<td>60.5</td>
<td>41.7</td>
<td>29.2</td>
<td>54.5</td>
<td>52.2</td>
</tr>
<tr>
<td>III-R</td>
<td>55.0</td>
<td>35.7</td>
<td>78.9</td>
<td>27.9</td>
<td>44.6</td>
<td>51.9</td>
</tr>
<tr>
<td>III-FS</td>
<td>20.7</td>
<td>39.1</td>
<td>87.0</td>
<td>51.9</td>
<td>50.0</td>
<td>59.3</td>
</tr>
<tr>
<td>III-I</td>
<td>-</td>
<td>54.4</td>
<td>63.6</td>
<td>41.3</td>
<td>55.9</td>
<td>31.8</td>
</tr>
</tbody>
</table>

Table 3: Scores (in %) for correct responses for vowels in word items put, bull sugar, sugar, food, moon, flute, juice, /ʃrʌθ/; belive, leave, weeds, trees, trees, please, /dʒiːl/, (lived, _lɪve, _piɡ), kids, milk, /θʌtʃ/; red, _hɛd, _bɛɡ, _lɛtər, _bɛd, _bɛdʀoʊm, /stɛɡ/, /θrɛbz/, /smeθ/, months, months, /brəndæ/ in different test types and across age groups.
Table 4: Scores (in %) for correct responses for vowels in word items (last, calf, glasses, bananas), half, bathroom, /dɔːt/, hammer, hamburger, have, cat, /ˈkeɪhə/, /ˈpɪkʃərər/, /klæθ/, (dogs, cost, lot), log, clock, socks, (slender, water), matter, winter, father, /ˈkeɪtəs/ worm, words, water, water, /kræz/ in different test types and across age groups.

<table>
<thead>
<tr>
<th></th>
<th>/ɑː/</th>
<th>/æ/</th>
<th>/ɒ/</th>
<th>/ɔː/</th>
<th>/ə:/</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-R</td>
<td>8.8</td>
<td>24.2</td>
<td>40.9</td>
<td>6.0</td>
<td>-</td>
</tr>
<tr>
<td>5-FS</td>
<td>20.3</td>
<td>34.6</td>
<td>56.5</td>
<td>21.4</td>
<td>-</td>
</tr>
<tr>
<td>5-I</td>
<td>2.7</td>
<td>55.0</td>
<td>-</td>
<td>8.3</td>
<td>-</td>
</tr>
<tr>
<td>7-R</td>
<td>7.9</td>
<td>35.1</td>
<td>51.0</td>
<td>5.7</td>
<td>0.0</td>
</tr>
<tr>
<td>7-FS</td>
<td>28.6</td>
<td>45.7</td>
<td>58.3</td>
<td>9.3</td>
<td>2.3</td>
</tr>
<tr>
<td>7-I</td>
<td>2.4</td>
<td>63.0</td>
<td>-</td>
<td>5.2</td>
<td>-</td>
</tr>
<tr>
<td>I-R</td>
<td>6.5</td>
<td>25.0</td>
<td>53.6</td>
<td>3.8</td>
<td>0.0</td>
</tr>
<tr>
<td>I-FS</td>
<td>23.5</td>
<td>51.5</td>
<td>62.5</td>
<td>17.9</td>
<td>3.6</td>
</tr>
<tr>
<td>I-I</td>
<td>5.0</td>
<td>50.0</td>
<td>-</td>
<td>2.7</td>
<td>-</td>
</tr>
<tr>
<td>III-R</td>
<td>10.1</td>
<td>32.1</td>
<td>56.3</td>
<td>3.6</td>
<td>0.0</td>
</tr>
<tr>
<td>III-FS</td>
<td>20.0</td>
<td>38.7</td>
<td>60.0</td>
<td>29.6</td>
<td>3.7</td>
</tr>
<tr>
<td>III-I</td>
<td>0.0</td>
<td>45.5</td>
<td>-</td>
<td>6.3</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4 shows the results of the analysis of correct responses for the pronunciation of English sounds by Slovene learners. As mentioned, the research was primarily designed to elicit ‘erroneous’ responses, above all those which would prove L1 dialect interference (the test items were chosen accordingly) and also to observe responses coming as a result of interference of other varieties of English (e.g. General American\(^1\), since the exposure to it in Slovenia is abundant due to the multitude of mass media and Internet connections, and it is despite the formal ESL

\(^1\) RP was here taken as the basic model for ‘correct’ results since ESL instruction in Slovenia is based on it. When it comes to responses obviously influenced by the most common different varieties of English (i.e. British and American English), the British variety was for the sake of clarity considered ‘correct’ and the American variants ‘deviant’, even though both ‘correct’ and ‘deviant' responses were in the study presented as the interlanguage or interphonology of the Slovene English spoken by Slovene learners. If results seemed questionable due to this classification, it should be noted that it does not affect the results of this part of the study, since both varieties were proportional in all test types, and it is the relation between test types that is the point of interest here.
instruction in RP bound to show in the pronunciation of Slovene learners) and those which can be characterized as “psychological” speech errors (epentheses, substitution, anticipation, etc.). The above mentioned responses, however, were not considered an opposition to ‘correct’ responses, but were rather treated together with them as ‘errors’ side-by-side with ‘non-errors’ forming individual stages in the interlanguage of Slovene learners of English.

Although the main focus of the PES was on ‘errors’, the ‘correct’ responses or ‘non-errors’ were not neglected. As already mentioned, they were incorporated into the tables and discussed in the context of orthographic interference in the Slovene English interlanguage, but it was only with the justification of test type that they gained greater attention. This is not something new. Some linguists have gone to great lengths to identify influence of test types on learner speech, as will be shown later.

Since Labov in the 1960s, linguists have shown interest in the influence, if any, test type might have on results obtained. Different linguists arrived at different and sometimes inconclusive results in different test situations. The present paper will introduce the views and results of three linguists considering test-type impact, namely Labov, Dickerson and Flege. Labov (1966) dealt with L1 speakers of English in New York, and Dickerson and Flege with L2 adult learners. The results of Labov’s study proved that the highest amount of ‘correct’ responses was obtained in the reading test type which according to him contained the larger amount of standard markers and required the largest amount of attention on the side of the speaker. He observed that “attention varies systematically with different verbal tasks” (In Wenk 1979: 211). Attention increases with the increase of standard markers and with it correctness. Rapid speech is said to contain fewest standard markers, but in response to interview questions, speakers use more educated variants. Still more formal variants appear in passage reading, and even more in word-list reading. Dickerson (1977) made similar observations in adult second language. He found that learners of English as a foreign language have the smallest number of target or target-like or correct variants in their unpremeditated or free speech, progressively more in general reading, and still more when reading word lists or minimal pairs. Flege, on the other hand, got inconclusive results, especially when it came to pronunciation. He tackled the problem from a different angle and the opposite point of view. He tested the hypothesis that “L2 learners produce L2 sounds with decreasing authenticity as the attentional demands of the speaking task increases” (Flege 1987: 30), meaning that they would get fewer target-like or correct responses in reading (where attentional
demands are higher) than in free speech (where the respondents become more relaxed) which was an opposing view to the two mentioned above. He concludes that different sounds may behave differently when used in reading or in free speech and adds another factor to consider, namely that different results could be expected if younger L2 learners were observed rather than adult L2 speakers (Flege 1987: 31).

PES provided results which corresponded to none of the above mentioned theories. Deviation from Labov’s assumptions could be explained with the fact that he dealt with L1 adult speakers, while PES included young L2 learners. Dickerson’s claims are supported by the fact that although his subjects are speakers of ESL, they are adults. Neither Labov nor Dickerson mention the level of difficulty of the test types. Most certainly we might assume that the ratio of ‘correct’ results across test types may change as the tasks (either reading or free speech) increase in difficulty. It could be assumed that reading becomes “easier” with the raising of level of difficulty (due to the visual support) and free speech more difficult due to the lack of such support (if it is not altogether non-existent – if the respondent fails to produce the word or phrase in question). Flege might have got conflicting results, namely some sounds performed better in reading tests and others in free speech, in adult ESL learners due to relative but not extreme difficulty of the texts.

In the case of the PES we were due to the pseudo-longitudinal nature of it forced to adjust the level of test type (be it written, free speech or imitation) to the youngest respondents, namely to the fifth-graders. It was probably due to this familiarity with the meaning of the test items that the ‘correct’ responses scored highest in free speech rather than reading. One of the reasons for this might be that we were mainly interested in whether the respondent was capable of producing a certain sound in a certain structural position rather than determining whether s/he knew the (same) word in reading or in free speech.

---

2 It should be noted that all the words and sentences elicited from the respondents, be it free speech (including translation tasks which enable more economical eliciting of desired sounds or reading were adjusted to the level of the current L2 instruction for the respondent and it was assumed that the respondents should have no difficulty identifying them.
5. Interpretation of the results

Tables 1-4 show the results of the comparison between correct responses across test types (R - *reading*, FS – *free speech* and I – *imitation*) and across age groups for most of the sounds in the English sound system. Tables 1-2 represent the consonant sounds (mainly the obstruents, because the differences in results were the greatest here), while Tables 3-4 display the results for all the monophthongal vowel sounds.

While the results for the vowel sounds are consistent with the observation that *free speech* provides more ‘correct’ results than *reading*, the results for consonant sounds are not as consistent. An observation can also be made in the case of consonants – both types of voiceless consonants, plosives (Table 1) and fricatives (Table 2), varied in the results, in some cases *free speech* prevailed, while in others *reading* did. In the case of the voiceless plosives this could be explained by the fact that these sounds were in all three test types observed particularly in the structural context “initial position in stressed syllable before a vowel”, which in English requires aspiration. Typically, Slovene learners of English will produce erroneous responses at this point. However, introducing test types such as word lists might provide support for the respondent in terms of the energy invested into the reading of individual words in the list, which gives this test type an advantage compared to *free speech* where the context provides no such conditions for the onset of the initial voiceless (fortis) consonant.

The lenis consonants testify to a higher number of ‘correct’ results for *free speech* rather than *reading* (which contradicts the findings made by Labov and Dickerson, but confirms the assumptions made by Flege relating to earlier stages of foreign language learning.

The deviation of the results in this paper compared to those obtained by Labov and Dickerson may not have only to do with degree of attention paid to different test types, but also with the orthography of L1 and its relationship to the one in the L2. This relates to an observation made by the present author in another study in both English and Slovene in the beginner stage and in later stages of development. The respondents in the study were required also to read complicated structures such as tongue twisters. Although it was expected that the younger students (the fifth-graders) would produce more errors in pronunciation than the older ones (seventh-graders), the situation was the reverse. The fifth-graders produced fewer errors in reading than the seventh-graders although their level of proficiency was not as high as the one of the seventh-graders. It might be that they found normal sentences as difficult to pronounce as
tongue twisters and they paid the maximum amount of attention to both, which was not the case when we observe the seventh-graders. The study of the pronunciation of English in Slovenia (PES) mainly involved respondents who would in proficiency better fit the seventh-graders in the mentioned study (more fluency and less attention) rather than the fifth-graders (who more readily fit the formulation more attention than fluency).

The maximum attention of the fifth-graders and the better pronunciation or more ‘correct’ results in the category of free speech might also be attributed to the fact that the Slovene writing system is according to the ODH (Katz 1992: Orthographic Depth Hypothesis) one of ‘shallow orthography’ as opposed to English which is undoubtedly ‘deep’. Slovene ESL learners thus not only have to master the English writing system, which is already in itself extremely complex (infrequent letter-to-sound correspondence), but are also forced to overcome the obstacle called L1 interference of its ‘shallow orthography’ (general letter-to-sound correspondence), needing to grasp the concept of shifting from ‘shallow’ to ‘deep’ before entirely mastering the complexities of reading in English. On the basis of the above mentioned it no longer seems strange that free speech scored higher in correct results than reading.

6. The Orthographic Depth Hypothesis (ODH) and requirement for ‘parity’ as theoretical support of free speech over reading

The Orthographic Depth Hypothesis (‘deep’ vs. ‘shallow’ orthography) or ODH was introduced by Seidenberg (1992:102). As mentioned, Slovene is considered a phonographic language which has a strong correspondence between graph and phone. With only a few exceptions of vocalized “l” and “v” before consonants and a pause, words can be spelt (almost) in terms of phonemes. The English spelling system, on the other hand, differs considerably from the pronunciation of individual phonemes the letters represent. Garman even describes English spelling as being “compositionally phonographic only postlexically; in other words, the grapheme-phoneme mappings become apparent after the word in question has been identified, and are not an adequate basis on which to achieve identification.” (Garman 1990:38) Proof for this would be the many-to-many mappings between sounds and letters, e.g. in ration, ratio, patio, Thames… Also, differences in spelling between homophones such as too vs. to vs. two, for vs. four, be vs. bee, etc are considered visual rather than phonological (Garman 1990: 40), which would imply that the unit of
English orthography is the word and is perceived similarly as e.g. Chinese pictographs, whereas Slovene is a purely alphabetic system.

Psycholinguists in the field share the opinion that in all orthographies, even in Chinese, phonology plays at least some role in word recognition. Liberman (1992: 170-171) takes an entirely different route and claims that letter-to-sound correspondence is not that simple or straightforward. He argues that speech and reading/writing are not equally natural, that is, they are not psychologically equivalent. He and Mattingly introduce the so-called requirement for “parity”, i.e. something that reading and writing do require and speech does not. While letters are mediators in the case of reading/writing, sound signals are not so in the case of speech. What we produce are not speech sounds in the sense of phones or allophones, when it comes to phonetic shapes, but coordinated movements of gestures manifested through time (cf. also Browman and Goldstein 1990). Liberman (1992: 174) adds that (t)here is no need, as there is in the conventional view [note: which claims that phones are mediators in speech as letters are in reading] for the cognitive translation from an initial auditory representation [note: e.g. listening to someone else speak or listening to one’s own pronunciation of the utterance, then recognizing it, then reproducing it], simply because there is no auditory representation.” A mediating auditory representation does not seem to exist in speech. What seems to exist in its place is the real thing, i.e. the coordinated movements of gestures producing the stream of speech. Viewed from this point it would seem that the read-and-write process is least complicated in those orthographies which actually have a more pronounced letter-to-sound correspondence. To quote Katz (1992:2) “A common proposal has been that the more closely an orthography conforms to the alphabetic principle, the more efficiently phonological representations will mediate between print and lexicon.” This would also mean that learners used to this letter-to-sound correspondence in their own language would have additional problems in dealing with writing systems in which the letter-to-sound correspondence is not so pronounced (e.g. Slovene learners of English). Such learners will find reading in English, for example, additionally difficult.

7. Conclusion

Studies in applied linguistics have for decades employed different methods of collecting data. At linguistic levels higher than phonetics and phonology, the test type proved to be an important issue as it was obvious that if the investigator wanted to obtain results in comprehension, lexical
issues, grammatical tasks, reading as the sole method would not give feasible results. Other methods, frequently indirect ones, were thus needed to elicit the mentioned types of knowledge. Phonetics and phonology, however, present a different problem. Not having to consider grammatical and semantic issues in these fields, reading has lent itself as a perfect solution for getting the respondents to pronounce the desired utterances, since the written form of the words usually gave away the meaning rather than the pronunciation and it seemed that nothing was lost by employing reading as the method of eliciting responses. Not only that nothing was lost, time was gained as the respondents did not “beat around the bush” but went directly to the issues in question. Consequently, the conditions of obtaining data were more uniform and thus better suited for comparison.

Although the question of feasibility/reliability of the results obtained in such a manner was not questioned, some phonological studies, as already mentioned, did attempt to include other methods of eliciting results, i.e. test types involving reading (e.g. Eckman) and others involving translation, free speech on the basis of a specific topic of conversation, and imitation tests (e.g. Wieden/Nemser), but designed in such a manner that they could not account for the differences coming as a result of different test types (especially between reading and free speech ones) unless the study was geared toward observing the influence of different test types on pronunciation. Moreover, as also mentioned in this paper, studies (Labov, Dickerson, Flege) were carried out in order to determine the influence of test type on pronunciation, each arriving at different, even conflicting results.

Admittedly, the present study, namely PES, was not primarily designed to observe the influence of different test types on the pronunciation of English, but it did contain elements enabling the gaining of insight into the issue. The results obtained were counter to those obtained by any of the authors mentioned in the paper (i.e. Labov, Dickerson and Flege). Even though it would be difficult to generalize such a complex issue as the influence of test types on results, which rely largely on factors such as uniform speaking conditions, attention span of the informants, level of difficulty of the tests, the way they are administered, to whom they are administered, whether it is a matter of first language or second language, etc., several have tried to do so (Labov claiming that reading will score higher in ‘correct results’ than free speech based on standard markers – but he dealt with L1 adult speaker, Dickerson claiming the same as Labov – but he dealt with adult L2 speakers, Flege being undecided as to influence one way or the other, but not excluding the possibility that some sounds will be influenced and others not). The PES gave results which
contradicted all the above mentioned. But we could add that although we
are not familiar with the level of difficulty of the tests in the prior studies,
the level in the PES was such that almost all respondents were familiar
with the lexical items in the tests; they were mainly fluent communicators
in English, and last but not least, if not to similar studies in general, the
results of the PES could apply to research done in SLA where the first
language is one of shallow orthography and the second one of deep
orthography. If nothing else, it emphasizes the importance of free
(spontaneous) speech in assessing proficiency in L2 pronunciation. And
with technology allowing the storage of enormous amounts of recorded
data and easy access to it, choosing free/ spontaneous speech as the
method of obtaining phonological data in fieldwork not only becomes the
necessary but also the obvious choice.

References

Browman, C. P. and L. Goldstein. 1990. Tiers in Articulatory Phonology,
with Some Implications for Casual Speech. In Papers in Laboratory
Phonology I. Between the Grammar and the Physics of Speech, 341-
376. Cambridge: CUP.

Dickerson, W. B. 1977. Language Variation in Applied Linguistics. ITL
35: 43-66.

Eckman, F. 1987. The Reduction of Word-final Consonant Clusters in
Interlanguage. In Sound Patterns in Second Language Acquisition,

Flege, J. E. 1987. Effects of Equivalence Classification on the Production
of Foreign Language Speech Sounds. In Sound Patterns in Second
Dordrecht: Foris.


Maribor: Zora.

Katz, L. and R. Frost (eds.). 1992. Orthography, Phonology, Morphology,
B.V.


Orthography, phonology, morphology, and meaning, edited by R.
USING WEB TECHNOLOGIES IN L2 PHONOLOGICAL RESEARCH: METHODOLOGICAL ISSUES AND IMPLICATIONS

ANASTAZIJA KIRKOVA-NASKOVA AND DIMITAR TRAJANOV

Outline

This paper reports on the benefits and limitations of a web-based experiment designed for the purposes of a broader study investigating the English speech of Macedonian learners of English. The subjects of the study were first and second year students of English at Skopje University whose speech was recorded and then evaluated by trained phoneticians, native speakers of English, using a specially designed web application as a data gathering instrument. The results demonstrate the manifold benefits of the approach as well as the flexibility of its adaptation in applied linguistic research and L2 learning/teaching.

1. Introduction

The use of computer technology in applied linguistic research has become common practice. In the field of phonetics, phonology and second language phonological acquisition, computer assistance has shown advancement and yields remarkable results. Chun (2007) observes that various hardware devices and specialist software packages have been designed for the purposes of speech research in general—from acoustic research on segmentals and prosody to the creation of speech corpora (recording, archiving, tagging, transcribing) and the development of speech synthesis and speech recognition programs. Chun further notes the growing use of technology for phonological component manipulation in studies related to the perception and production of L2 speech, foreign accents, intelligibility, the comprehensibility and acceptability of spoken texts, L2 prosody, the development of listening and speaking competence,
and the use of technological tools and programs for L2 pronunciation instruction and learning.

Research practice with foreign accent ratings and segmental analysis studies has shown that the traditional methodology applied in conducting such experiments can be lengthy and wearisome for researchers and participants alike (Beddor and Gottfried 1995). For researchers, such methodology usually involves an extensive preparation stage for speech stimuli design. Participants, on the other hand, are typically expected to listen to repetitive recorded speech stimuli and to complete various identification/discrimination tasks, and sometimes to provide detailed phonetic analysis. In doing so, they may become impatient, bored or indifferent, thereby compromising the validity of the responses. To overcome these problems, a number of studies have incorporated computer-aided tasks and procedures (e.g., Escudero 2000). Such studies have typically been undertaken in highly controlled laboratory conditions, with expensive equipment and a large number of participants.

In addition to the expensive equipment and the length of the experiments involved, a potential shortcoming in phonetic studies that collect responses of segmental elements and global accent ratings is the specific listeners’ profile required for the experiment and their availability at the time of the study. Although untrained listeners are perfectly capable of making appropriate evaluations of non-native speech (Brennan and Brennan 1981; Anderson-Hsieh and Koehler 1988; Derwing, Munro and Wiebe 1998), many studies have found that experienced listeners, namely those with phonetic-phonological expertise, detect foreign-accentedness more reliably (Flege 1984; Thomson 1991). In fact, Flege (2002) argues that, despite the possibilities of gathering quantitative measurements of foreign-accented speech, it is the qualitative judgments of native speakers which remain “the golden standard”.

Only recently have the possibilities offered by the Internet begun to be explored. The Internet hosts a vast number of general linguistic questionnaires and surveys and quiz-based applications for practicing phonetic symbols and target sounds. To our knowledge, however, the only study to have employed a large-scale Internet survey for foreign accent evaluation by native speakers was that conducted by Rias van den Doel (2006). In this study, Van den Doel addresses the issues of the intelligibility and acceptability of Dutch pronunciation errors in sentences read by bilingual actors and evaluated by more than 500 native speakers of different Standard English varieties. The software used in the development of this survey is WWStim (Veenker 2003), a Perl CGI script for presenting
web-based questionnaires and experiments by using predefined sequences of template-based HTML pages.

This paper presents the development and design of a web application created as a user programme to gather native speakers’ responses to Macedonian-accented English speech. We begin by giving a brief overview of the aims of the study and the research context. We then turn to describing the application design and its implementation. In the final section of the paper a summary of the phonetic results is included, followed by a discussion of the advantages and some shortcomings of the approach as reported by the listeners and encountered by the researchers.

2. The study: Research context and choice of methodology

The pronunciation of English by Macedonian learners of English demonstrates systemic features that are predominately the result of interaction between the phonologies of these two very different languages. Siljanoski (1976) has made an attempt to compare the two systems employing contrastive analysis. His conclusions identify potential difficulties for Macedonian learners acquiring English pronunciation in the phonemic, allophonic and phonotactic domain (the prosodic level was not included in his study). Unfortunately, these conclusions have never been tested experimentally.

Another aspect that must be taken into consideration is the way native speakers perceive features of foreign accent in the pronunciation of non-native speakers. Munro and Derwing (1995, 1998, 1999, 2001), for instance, examine the relationship between foreign-accentedness, intelligibility and comprehensibility. Their findings reveal that heavily accented speech is indeed more difficult for native speakers to understand. However, even heavily accented speech is sometimes fully intelligible and comprehensible and does not hinder communication. Other factors which have been found to have the strongest impact on overall intelligibility and comprehensibility involve a combination of grammatical and pronunciation errors, with prosodic errors being perceived as potentially more detrimental than phonetic errors. It seems that the globalisation of the English language may have caused a shift in native speaker attitudes towards pronunciation deviating from the norm. Not all mispronunciations, then, are perceived by native speakers as being equally detrimental. Rather, the greatest intolerance on the part of native speakers is demonstrated with regard to pronunciation errors that impede communication and cause unintelligibility.
With this in mind, our study aimed to detect and describe those segmental sounds in Macedonian-accented English speech which native speakers of English most frequently perceive as deviating from their own standard variant of English. Our research additionally sought to ascertain whether native speakers of different English variants perceive the same segments as non-native speakers. During the preparation stage, it became evident that our research would require a specific target group of native speakers: English native speakers speaking different English varieties and with phonetic training. As no such high profile experts resided in Macedonia, we developed for the purposes of our study an experiment in which the English speech of Macedonian native speakers was evaluated by English native speakers with the use of a specially designed computer web application employed as a data gathering tool.

2.1. Participants

The participants who took part in the research undertaken for this study comprised: (1) Macedonian learners of English who provided speech samples of their English for evaluation; (2) a native speaker of English employed as a control speaker; and (3) native speakers of English recruited to serve as ‘raters’ in the evaluation process.

A total of 17 Macedonian native speakers were recorded. All were first and second year students (aged between 19 and 25) residing in Skopje and majoring in English Language and Literature at the University of Ss. Cyril & Methodius. Their level of English ranged from B2 (n=6) to C1 (n=11). None of the group had resided for any extended period in an English-speaking country.

One native speaker of British English was recruited to serve as the control speaker for the accent-rating section of the experiment. His speech was part of the experiment among the recordings of the Macedonian speakers to provide a standard by which to assess the reliability of the other native-speaker evaluators. The choice of only one native control speaker, and not more as is the practice with large-scale experimental research studies under laboratory conditions, was deliberate in view of the small scale of this study and the consequent risk—as reported in Angelovska and Hahn (2009)—of native speakers being rated as ‘near-native’, implying non-native, and thereby bringing into question the reliability of the native speakers as raters.

14 native speakers of English completed the experiment and rated the speech samples obtained from the Macedonian learners. Their age ranged from 28 to 71 years old (median 49). All had university degrees or
postgraduate qualifications (PhD=5, MA=6, BA=3) and phonetic-phonological expertise. Some had prior experience as raters (n=10). Most were either active or retired university lecturers (n=8), while the remainder included English teachers (n=4), a translator (n=1) and a language consultant/book author (n=1). As reported in the questionnaire, they spoke the following English variants: Southern British English (n=6), General American English (n=6), Irish English (n=1) and Canadian English (n=1). Most of them had never interacted with speakers of Macedonian origin, though two raters reported brief exposure to Macedonian and Macedonian-accented English.

The raters were mostly recruited via the websites of LINGUISTLIST http://www.linguistlist.org/ and the IATEFL Pronunciation Special Interest Group PronSIG http://uk.groups.yahoo.com/group/iatefl_pronsig/, or were directly approached via e-mail and/or personal acquaintance.

2.2. Stimulus materials

The recordings obtained from the Macedonian speakers and the control speaker were recorded in a soundproof booth using a Sound Forge 8.0 computer programme and a 20-channel 4-BuS Mixing Console Behringer MX2004A.

The speakers were each given a ‘free speech’ task in which they were asked to select one of four possible topics to speak about for 1–2 minutes. They were instructed to speak at their normal pace and preferably not to use words or phrases that might reveal their Macedonian origin. The recordings were then edited to an average approximate duration of 17–29 seconds (Jesney 2004) and then tested for authenticity.

The topics included in the free speech task were intentionally designed to be personal and with an emotional overtone in the hope of getting speakers to re-experience the events they described and thus focus more on the content of their speech than on the form of their delivery (Dowd 1984). This approach encouraged the speakers to demonstrate a natural flow of speech and eliminated the possibility of their rehearsing set speeches (Munro and Derwing 1994). As a result, their speech resembled a narrative style of spoken discourse typical of informal settings (Oyama 1976; Major 1986; Thompson 1991). Other considerations in the choice of free speech as an eliciting technique (in preference, that is, to reading word lists, sentences or paragraphs) included the desire to avoid the possibility of pronunciation errors arising due to lack of word familiarity or even negative transfer of their L1 reading skills combined with orthographic interference (Munro 2008).
3. Web application

The web application employed in this study is an example of an Internet-based research method that works well with any Internet browser available, such as Internet Explorer 7+, Mozilla Firefox 2+, Opera or Google Chrome. Table 1 details its technical specifications:

<table>
<thead>
<tr>
<th>Technology</th>
<th>Microsoft ASP.NET 3.5, Adobe Flash</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programming Language</td>
<td>C#</td>
</tr>
<tr>
<td>Development Platform</td>
<td>Microsoft Visual Studio 2008, Adobe Flash CS3</td>
</tr>
<tr>
<td>Operating System</td>
<td>Microsoft Windows Vista</td>
</tr>
<tr>
<td>Database</td>
<td>Microsoft SQL Server 2005</td>
</tr>
</tbody>
</table>

The application was developed as a user programme that would be able to display predefined sets of questions, to store data (i.e. keep track of the answers checked by each of its users) and to create an easily accessible database for further analysis. In addition, the following components were incorporated:

- the experiment administrator was able to create accounts for the users (public access was not allowed)
- answers were chosen from check box lists or radio buttons
- free text boxes were added so that the listeners could make comments on speech phenomena they had heard and noticed
- security measures were undertaken so that users could not manipulate their answers once they had been submitted (they were not allowed to use the back button). They were, however, allowed to opt for a break during the experiment.

The application was organized in separate webpages, as follows:

1) an introduction with instructions (Figure 1)
2) a user account window
3) a participation agreement
4) a questionnaire for personal background details
5) experiment questions (Q1–Q4) with audio files and comment boxes repeated for every speaker (Figure 2)
6) comment boxes for users in order to:
   a) insert their impressions of Macedonian-accented English
   b) evaluate the experiment format (Figure 3).
Figure 1: Experiment start page

The web page with the questions and audio files was the core of the experiment. Two types of data were collected: a) quantitative data on the frequency of phonetic segment variables and foreign accent rating—Q1: consonantal variables; Q2: vocalic variables; Q3: global foreign accent ratings on a 5-point scale (adapted from Bongaerts et al. 1997: 456); Q4: general variables for foreign accent evaluations)—and b) qualitative data from open-ended questions with comment boxes. The variables used in experiment questions 1 and 2 were selected on the basis of our teaching experience and the most frequent mispronunciations reported in the literature consulted (Weinberger 1998; Siljanoski 1976, 1993; Dimovska 1980). The variables in Question 4 presented a selection of factors tested in other studies (Munro and Derwing 1998; McDermott 1986).
The administrator first created accounts for each rater and then sent a username and password via e-mail. The raters were expected to log into the system at http://e-tech.feit.ukim.edu.mk/phonetics/. They were then redirected to the next webpage to complete the experiment. The raters were asked to listen to each individual speaker’s audio file and answer the questions related to each file as instructed. The procedure was repeated for every speaker. If a rater did not complete the whole experiment in one sitting, they could log in again and continue where they had stopped as the application kept track of the users’ answers. The total duration of the experiment was 60–90 minutes (as reported by listeners who completed the pilot version of the experiment in one sitting).

The development of this application enabled the creation of a small-scale corpus of listeners’ responses. This allowed for access to pre-stored data that could subsequently be grouped for faster analysis. On completion of the experiment, the data was extracted in a readable format using Microsoft Excel 2007. Additional statistical analysis for the calculation of the frequency of checked variables was conducted using the statistical
package SPSS 16. The written comments were compared and generalisations inferred.

**Figure 3.** Impressions and Comments page

![Impressions and Comments page](image)

### 4. Results and discussion

As the main emphasis of this paper is on the experimental design of the study, the phonetic results will only be briefly explained.

The results analysis is divided into two parts. First, the findings related to the frequency distribution of consonantal and vocalic segments in the English speech of Macedonian learners evaluated as non-native by English native speakers were interpreted. Thus, the feedback from 12 native speakers—6 British native speakers (BNS) and 6 American native speakers (ANS)—was considered as two separate groups. These responses were then compared with the responses of the Canadian and Irish native speakers (given that one speaker cannot be regarded as a group). The second part of the analysis interprets the responses provided in the comment boxes from all 14 native speakers as one group.
The data for the consonantal variables is presented in Table 2 and for the vocalic variables in Table 3.

Table 2 shows that the highest frequency of perceived sound deviations is indicated for final obstruent devoicing by both groups of raters. A similar number of responses by both BNS and ANS is noted for the pronunciation of dental fricatives pronounced as dental plosives, with a higher percentage for /ð/>/d/ and lower for /θ/>/t/. The rest of the variables were less frequently checked by both groups of native speakers. Nevertheless, slight parallels do occur with: the English glottal fricative pronounced as Macedonian velar fricative /h/>/ɯ; velar plosives accompanying the pronunciation of the nasal velar /ŋ/>/ŋ, ŋ; postalveolar /ʌ/ pronounced as an alveolar tap or trill; and bilabial approximants pronounced as labio-dental fricatives /w/>/v/. It seems that the American raters were more sensitive to the non-aspiration of voiceless plosives /p, t, k/ and the dentalization of English alveolars /t, d/. The British raters, on the other hand, noticed the insertion of /ʌ/ in every word position—as expected.

**Table 2. Frequency distribution of consonantal variables**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>final obstruent devoicing</th>
<th>non aspiration</th>
<th>dentalization</th>
<th>θ&gt; t</th>
<th>ź&gt;d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>BNS</td>
<td>102</td>
<td>52 50,98</td>
<td>6</td>
<td>5,88</td>
<td>3</td>
<td>2,94</td>
</tr>
<tr>
<td>ANS</td>
<td>102</td>
<td>58 50,86</td>
<td>16</td>
<td>15,69</td>
<td>9</td>
<td>8,82</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>w&gt;v</td>
<td>ȗ&gt;r</td>
<td>h&gt;x</td>
<td>ŋ&gt;ŋ / ŋk</td>
<td>ȗ insertion</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>BNS</td>
<td>102</td>
<td>2  1,96</td>
<td>6</td>
<td>5,88</td>
<td>10</td>
<td>9,80</td>
</tr>
<tr>
<td>ANS</td>
<td>102</td>
<td>3  2,94</td>
<td>6</td>
<td>5,88</td>
<td>14</td>
<td>13,73</td>
</tr>
</tbody>
</table>

The frequency results for the vocalic variables in Table 3 indicate a higher degree of tolerance among native speakers. It seems that both BNS and ANS predominantly perceive vowel shortening in the pronunciation of vowels by Macedonian speakers. The lower percentage of checked responses for the other vocalic variables leads to the conclusion that vocalic mispronunciations are quite acceptable. Differences between the
groups were noted with changes in vowel quality, that is /æ/ and /ʌ/ pronounced as /e/ or /a/ respectively, which were observed more by ANS than BNS.

Table 3. Frequency distribution of vocalic variables

<table>
<thead>
<tr>
<th></th>
<th>S 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNS</td>
<td>102</td>
<td>47</td>
<td>46.08</td>
<td>4</td>
<td>3.92</td>
<td>3</td>
</tr>
<tr>
<td>ANS</td>
<td>102</td>
<td>41</td>
<td>40.20</td>
<td>16</td>
<td>15.69</td>
<td>1</td>
</tr>
</tbody>
</table>

When a comparison is made between the summarised results from the British and American raters and the responses from the Canadian rater and the Irish rater, a certain degree of consistency is noted. These two raters also perceived consonant deviations as most striking, particularly with regard to final obstruent devoicing. Except for /æ/>/ɟ/, which was marked by both raters, other vocalic adaptations were mainly accepted. The Irish rater also pointed out vowel shortening.

In the comment boxes, the raters were asked to note any specific examples of words in which they noticed segmental deviations in pronunciation. In addition to words exemplifying some of the listed variables, the raters also pointed out other phonetic phenomena not included in the list: vowel raising, vowel shortening, diphthong levelling, medial consonant voicing, lateral velarization, overcorrection, regional variant preference (American vs. British). Although not part of the study focus, frequent comments related to prosodic deviations and connected speech emerged spontaneously.

Based on the gathered data, we were able to pinpoint the pronunciation characteristics of a typical Macedonian learner of English (a representative of the study sample):

- **Consonants** are clearly pronounced but completely devoiced in final position (obstruents). New consonantal sounds are substituted
with their closest Macedonian equivalent (ð/>/d/; /θ/>/t/; /w/>/v/), whereas similar sounds are adapted (place or manner of articulation, e.g. h/>/x/; /t, d/>/t , q/; /η/>/ηk, ηg/). /r/ is realised as a tap and pronounced in all positions; /l/ is velarised and pronounced rather hard; and voiceless aspirated plosives [pʰ, tʰ, kʰ] are either weakly aspirated or not at all.

- **Vowels** are systematically adapted to the Macedonian five-member vocalic system. Long vowels are poorly distinguished and shortened (i/>/i; /a:/ > /a/, /ɛ:/ > /o/, /ɔ:/ > /u/, /ɔ:/ > /o/). All vowels are either raised or lowered (iɛ/>/ɛ; /s/>/a; /ɔ/>/a; /u/>/u; /o/>/o). Diphthongs are monophthongised.

- Aspects related to connected speech, stress, and intonation include the pronunciation of both stressed and unstressed syllables with equal length, sporadic use of weak forms, recurrent use of short tone units and rising intonation, and frequent pauses in unexpected places.

The responses obtained from the comment box regarding the experiment format suggest that the experiment was properly administered given the satisfactory number of native listeners who responded to the project advertisement and their successful completion of the experiment. The raters highlighted the clarity of the instructions, the high quality of the recordings in the audio files and the well-defined and user-friendly experiment as a whole. Some of them complained of the length of the experiment; others observed the need for a pause button in the audio file icon and the inability to use phonemic symbols at times.

The use of the Internet as a medium also proved advantageous in making the application accessible from anywhere in the world. It addressed a wide audience, with respondents from the USA, Canada, the UK, Ireland, and New Zealand. As the procedure was pre-programmed and could be administered in various locations at the same time, there was no need for the researcher to supervise experiment completion directly. The users were allowed to conduct the experiment according to their own schedule. This proved to be time-saving as data was quickly and efficiently gathered, allowing the researchers to receive immediate results and having a database instantly created and regularly updated. On completion of the experiment, all data was easily extracted and managed (no need for manual data input). In future, the application can be easily adapted for the requirements of similar research studies due to its flexible software design.
From a methodological point of view, some limitations were experienced with regard to the choice of subjects, the choice of the speech elicitation technique and the subjectivity of the ratings. The Macedonian native speakers were a homogeneous subject group, thus rendering results more indicative than significant (some of the features which appeared, for instance, are typical of Skopje dialect). Although they produced authentic/spontaneous speech featuring typical mispronunciations (which was very constructive for the outcomes), the recordings revealed that not all expected phonetic/phonological structures emerged naturally—as pointed out by Munro (2008)—because there was no guarantee that every segmental variable on the list would be mentioned by all subjects at least once in their speech samples, which directly influenced the frequency of the raters’ responses. In addition, considering our subjects’ proficiency level, various grammatical errors were produced which were perceived by the raters and noted in the comment boxes. A facilitated free speech task should thus perhaps be adopted in future research. Such a task might involve a choice of several topics and a set of phrases with segments under investigation for the subject to include in his/her speech. However, to preserve speech spontaneity, care should be taken not to overburden the phrases with too many difficult elements and to disallow subjects excessive preparation time. As a final point, it should be mentioned that the accented speech produced by the Macedonian speakers was treated as a one-dimensional phenomenon, thus ratings for intelligibility and comprehensibility were not measured. Although a high level of inter-rater reliability was confirmed, a certain amount of subjectivity was expected, especially bearing in mind that the raters’ first impressions could not be accounted for.

5. Conclusion

The methodology developed for our research is an example of how a modern approach can be applied when conducting phonetic experiments. There is room for its improvement and adaptation not only for research purposes but also in the area of language teaching, learning and assessment.

Such an approach, for instance, would be ideal for the promotion of research in less developed countries, as is the case with Macedonia, where people and resources are always limited and insufficient. Furthermore, it can be modified as part of learning management software where teacher-student interaction is preferred. Students can record themselves and upload their speech as an audio file. The teacher can mark their mispronunciations,
give immediate feedback, and monitor their progress. Alternatively, as part of their exams, students could be required to record themselves and upload the file for the teacher to assess their pronunciation according to a set of predefined phonetic items that are expected to be acquired throughout the academic year.

**Acknowledgements**

We are extremely grateful to Vlatko Apostolski and Goran Bakraceski from the Faculty of Electrical Engineering and Information Technologies, Skopje University, for their technical assistance and design of the web application.

**References**


Learning from Students’ Errors: English Phonetics Theory Exam

Rastislav Šuštarsič

Outline

This paper focuses on one part of the oral exam in English phonetics, namely, the ‘English Phonetics Theory’ part, which basically concerns testing students’ knowledge of the English (Standard British and American) sound systems (vowels and consonants), along with various word-internal allophonic realizations of vowels and consonants (aspiration, glottalization, voicing and devoicing, vowel duration, assimilation, etc.). I usually conduct this part of the oral examination by providing a short, simple English word such as crowds, and the student is asked to describe the sounds as they occur in that particular word. While a number of students show the expected level of knowledge and understanding of the basic features and processes occurring in the pronunciation of sounds in a particular phonetic environment, others seem to have considerable problems with regard to remembering, understanding and expressing the information I try to provide in my lectures and practical classes. It is from their (sometimes rather amusing) answers that I often get very important feedback on my teaching, in particular with regard to how certain explanations may be confusing or unclear for a number of students. Some of the observed students’ errors are at a level of understanding the basic notion of ‘position’ of a sound in a word, others concern somewhat more demanding understanding of e.g. the ‘influence’ of the neighbouring sounds on the sound in question.

1. Introduction

The paper focuses on one part of the oral English Phonetics exam for Slovene students of English (English Department and Department of Translation, Faculty of Arts, University of Ljubljana), namely the English Phonetics Theory part. This basically concerns testing students’ knowledge
of the English (Standard British and American) sound systems (vowels and consonants), along with various word-internal allophonic realizations of vowels and consonants (aspiration, glottalization, voicing and devoicing, vowel duration, assimilation, elision, etc.).

The examination is conducted by providing a short, simple English word such as *crowds*, and the student is asked to describe all the sounds as they occur in that particular word.

While a number of students show the expected level of knowledge and understanding of the basic features and processes occurring in the pronunciation of sounds in a particular phonetic environment, others seem to have considerable problems with regard to remembering, understanding and expressing the information provided in the lectures and practical classes. It is from their (sometimes rather amusing) answers that I can often get very important feedback on the effectiveness (or lack of effectiveness rather) of my teaching, in particular with regard to how certain descriptions and explanations may be confusing or unclear for a number of students. Some of the observed students’ errors, as specified and discussed below, are at a level of understanding the basic notion of ‘position’ of a sound in a word, others concern somewhat more demanding understanding of e.g. the ‘influence’ of the neighbouring sounds on the sound in question, etc. I have tried to classify the different problems I have identified, and obtained the ten types of error presented and discussed below.

2. Analysis: Ten types of error

Type 1: Not Connecting Notions (or Place and Manner of Articulation)

Students can find definitions for individual manners and places of articulation in a number of sources. In our case, they usually refer to either Gimson and Cruttenden 2008 or Collins et al. 2002. The first one gives, for example, the following definition of the term ‘plosive’:

‘A complete closure at some point in the vocal tract, behind which the air pressure builds up and can be released explosively (…)’ (Gimson and Cruttenden 2008:28)

The second source provides a slightly simpler definition:

‘Air is stopped in the mouth and let out with explosion.’ (Collins et al. 2002:13)
Students try to memorize one of these definitions, along with those for different places of articulation. The problem, however, is that they often treat place and manner of articulation as completely unrelated pieces of information. Thus, a common interchange between the teacher (i.e. myself, marked T in the examples below) and the student (S) tends to be as follows:

*T:* Why do we call /p/ a bilabial plosive?
*S:* It is bilabial because we use the lips. It is a plosive because we make a closure.
*T:* Where do we make the closure?
*S:* In the mouth. No, in the larynx
*T:* No. You’ve just said we use the lips. So, what exactly do we do with the lips
*S:* We round them.

At this point I ask the student to ‘try and combine the information’ they have about the place with the one about the manner of articulation. If even this does not help, then I just have to tell them the expected answer and hope they will use this knowledge for the description of the next sound in the word.

**Type 2: Not Making Distinctions Between Notions**

Students often find it difficult to distinguish between terms which are similar either in spelling, pronunciation or both, e.g. affricate/fricative, alveolar/velar etc. When, for example, they are asked about the manner of articulation of the first sound in *choked*, and they say it is ‘a fricative’, I tell them to put down the word, because they might actually be saying ‘affricative’, which is almost perfectly correct. Of course, the noun that is more often used in phonetics books is ‘affricate’ but a grammatically fully correct answer would be ‘an affricative’ rather than just ‘affricative’; yet at oral exams many students simply forget about the articles. As for the distinction between terms like alveolar/velar, I try to make it clear by providing the Slovene equivalents as well as the more descriptive terms teeth ridge and soft palate, respectively.

Some of the other problems concern the distinctions between voiceless/silent (consonants). Thus, students will often claim that there is a ‘voiceless’ /p/ in *psalm* or a ‘voiceless’ /r/ in *dark*. In addition, once they have learned that e.g. /p/ is a ‘strong, voiceless’ plosive, they tend to draw an equation mark between the two terms, which sometimes leads to this funny dialogue:
T: Is /p/ strong or weak?
S: Strong.
T: What does that mean?
S: That it’s voiceless.
T: And what does voiceless mean?
S: That it’s strong!

Also vague knowledge e.g. about glottal as having to do with the throat, the larynx and the vocal folds, but not more than that, leads to confusing the glottal stop and the glottal fricative, aspiration and glottalization, etc. One student at the exam was even convinced that all voiced sounds are plosives, because of the closing and opening movements of the vocal folds.

Type 3: The Vicious Circle

The following exchanges show either that the student 'is avoiding the issue' or the teacher should reformulate the question. Otherwise, we enter a 'vicious circle’, similar to the one above concerning the distinction between strength of articulation and voicing.

EXAMPLE 1
T: How do you know that /r/ should be pronounced in ‘free’?
S: Because I pronounce it.
T: Well, why DO you pronounce it?
S: Because it’s there!

EXAMPLE 2
T: Why is /z/ in 'easy' voiced?
S: Because the vocal folds vibrate.
T: OK, but why do they vibrate?
S: Because it’s voiced!

In both cases, we could easily avoid this problem, of course, by asking more specifically about ‘the position/environment in which’ /r/ is pronounced and the one in which /z/ is voiced.

Type 4: Birds of a Feather (Refuse to) Flock Together?

When explaining various phonetic processes, we should perhaps more often make it clear that when sounds follow one another, they tend to become alike rather than even more different than when they are apart. In a number of cases, students seem to believe that the opposite is true, as if
the usual process were one of dissimilation rather than assimilation. Thus they tend to claim that a segment is, for example, fully voiced because the one before it or the one after it or both is/are voiceless.

**Type 5: Sounds and Letters Don’t Agree**

I am always amazed during the introductory practical phonetics classes at the number of students who (although they all pronounce correctly simple words like cat) fail to give the correct answer once they are asked about the sounds (rather than the letters) in such words. They say, for example that:

*The first sound of CAT is a /si:/.
The vowel in BLACK is /ei/, so it is a diphthong.
The first sound of WRITE is /w/. No, it’s <w>. No, it’s /u/.*

All these cases may not be of the same kind, of course. Some students, I discover with horror, are actually convinced that there is a /w/ in write, and try hard to demonstrate that they make a clear distinction between *write* and *right* in their pronunciation.

**Type 6: Distinguishing between Before and After (a Sound)**

Strange as it may seem, the position of sounds in a word (or syllable) may not be perceived correctly by the student. They often produce the following statements, which of course are all wrong, but which I used to misinterpret in some way:

/ɪ/ is devoiced when before a /p/.
/ɪ/ is devoiced when it precedes a /p/.
/ɪ/ is devoiced when followed by a /p/.

When I started asking students to provide examples for this ‘devoicing’, they often gave valid examples like place, please, plot etc., so it became clear to me that the problem was the notions of before/after, not the students’ phonetic knowledge. It seems, therefore, that some students conceive the sequence of sounds (or letters) the way we conceive a sequence of objects or people in real situations, referring to the first one as the one on the right edge of the line, and not the other way round. I have started wondering if my explanation concerning the position of word stress might be adding to this confusion, because I often ask students to proceed ‘from right to left’ when trying to predict the stressed syllable of a polysyllabic word.
Type 7: Guessing: If it’s not Black, it Has to be White

One of the problems of oral exams is that (similarly to yes/no questions) we often ask questions about ‘binary’ features, i.e. features which are either ‘positive’ or ‘negative’. This leads to students’ guessing and their belief that they can easily get away with a quick reversal of the following kind:

S: /t/ is voiced.
T: Are you sure?
S: No, it’s voiceless. Of course.

A ‘naughty’ teacher will ask Are you sure? again perhaps, but this may not be the best way of dealing with this issue. It might be better not to question the ‘truth value’ of such statements (which is pretty much the same as telling the student they are wrong) but to ask them to explain the notion of ‘voicing’ (or some other binary feature) so that at some point they realize themselves that the answer was wrong. We could try and deal in the same way with a number of other (at least seemingly) binary concepts such as plural allomorph /s/ versus plural allomorph /z/, obstruent/sonorant, phoneme/allophone, fully voiced/devoiced, aspirated/unaspirated, etc. Some of these are actually not necessarily binary, so that students are unpleasantly surprised in situations like the following:

T: What kind of vowel is the one in ‘bird’, concerning its position in the vowel diagram?
S: It’s front.
T: No.
S: No, sorry, it’s back.
T: No.
S: ???

And the same happens with regard to ‘degree of opening’:

T: What kind of vowel is the one in ‘bird’, concerning its degree of opening?
S: It’s close.
T: No.
S: No, sorry, it’s open.
T: No.
S: ???
Type 8: Mispronunciations as Feedback on (Poor) Attendance

In the context of English Phonetics, one of the aims of our lectures (not to mention the practical classes) is, of course, to enable students to hear (one of the) the correct (i.e. standard) pronunciation(s) of a number of words, including the most important phonetic terms, such as vowel, consonant, tongue, plosive, alveolar, velar, glottal, etc. A student mispronouncing the words that the teacher has probably used dozens of times therefore reveals immediately (very often with the very first sentence they produce at the oral exam) their poor attendance in the classes.

Some of the most common pronunciation errors of my students are consonant, pronounced with a /z/ (an obvious influence of the corresponding Slovene loan word konzonant), plosive, pronounced with a monophthong), and interestingly the other way round, glottal (pronounced with a diphthong). I have tried to explain elsewhere (Sustarsic 2009) why in the tip of the tongue (in addition to a ‘spelling pronunciation’ of tongue with a back rounded vowel), a number of students pronounce a dental fricative for the first sound of all three words, not just the article. The word palate tends to be stressed and pronounced in such a way that it rhymes with inflate, and (notoriously) the first sound of the word vowel sounds just like the one in wow; this, however, is so common that it cannot be blamed on the lack of attendance, since also a number of very conscientious students find it difficult to pronounce the word with a labio-dental rather than bilabial consonant.

Type 9: Learning by Heart (from Unedited Notes?)

Missing classes necessarily leads to learning from various written sources, sometimes simply from a colleague’s scribbled notes, which will explain why some students will shamelessly try to reproduce whatever they have managed to decipher and memorize from them, without really thinking about the meaning of the weird expressions they have come across. Students in this way ‘invent’ meaningless words and phrases such as focal (i.e. vocal) folds, alveolar bridge (i.e. ridge), etc. Asking the student to try and explain what these expressions mean is often a waste of time, although it sometimes leads to amusingly bizarre explanations. The most common error of this type is the word rare (for rear) in the ‘rare’ part of the teeth ridge. Sometimes, I’ve recently realized, students know what they are talking about but only mispronounce the word (perhaps following the pattern of words like bear), but mostly it’s just learning a misspelt word by heart without ever checking its meaning. The quickest
way to see whether the student has some knowledge and/or understanding about what they have learned is to ask for a Slovene translation of whatever they are saying.

**Type 10: Phonetics and Logic**

The last type of error is closely connected with the previous one. When the answer makes no sense, it can hardly be argued about, as for example when you hear this:

_A closure is made with the teeth ridge on the soft palate._

This is just like saying that you can ‘scratch your back with your stomach’, and I’ve had to think very hard to provide an adequately silly comparison.

When this kind of answer does make some sense, however, it is often illogical, like the following:

_A closure is made with the tip of the tongue on the soft palate._

Faced with such claims, I try to demonstrate how such articulations might sound. Some other ‘creative errors’ of this kind are the following:

_The sound /r/ is much softer in GA than in RP._

_For /p/, the lips are rounded._

_We need a lot of energy to pronounce /t/._

Again, for this last claim, I try to show a sense of humour, and strongly recommend the student in question to avoid pronouncing too many ‘strong’ sounds or else they will soon be terribly exhausted. More seriously, however, I start wondering where they get this information. Most probably by ‘slightly’ changing my own unwisely formulated statement that fortis means strong but NOT in the sense of using a lot more energy etc. The phrasing of Gimson and Cruttenden, with the words energy and effort, may also be partly responsible for these overstatements:

_/p,t,k/ tend to be pronounced with more muscular energy and a stronger breath effort (...) (Gimson and Cruttenden 2008:159)_

**3. Conclusion**

I have tried to provide a brief analysis of some of the problems our students seem to have with ‘phonetic theory’, as manifested in the course of my oral exams in English Phonetics. The main aim of this analysis is
not, of course, to ridicule students’ performance at the exam but to try and find the main causes or reasons for their (often relatively poor) performance. I have also tried to suggest a few improvements in the process of teaching, which hopefully will be of some use also to other teachers of phonetics, either in terms of practical classes or lectures in phonetic/phonological theory. Perhaps the most useful activity in this respect, at least with regard to my own students, would simply be to give them a short presentation of the ten ‘types of error’ discussed above, so that they know what to look for and what to avoid in their process of preparing for the oral exam.

References


ENGLISH PRONUNCIATION MODELS AND TERTIARY-LEVEL STUDENTS: A BULGARIAN PERSPECTIVE

SNEZHINA DIMITROVA AND TSVETANKA CHERNOGOROVA

Outline

At the beginning of the 21st century, when English has undoubtedly established itself as the language of international communication, teaching English pronunciation to foreign learners at all levels of proficiency involves, among other things, making the important decision what native model, if any, the teacher should adhere to in the classroom. In the early years of the new millennium, the Lingua Franca Core was put forward as an alternative to traditional pronunciation models such as Received Pronunciation and General American (Jenkins 2000). However, scholars have argued that the teaching of a simplified form of English pronunciation could result in serious comprehension problems, since most non-native speakers are also listeners who have to cope with native-accented English (Trudgill 2005 and studies referred to therein). Thus, opting for a native-speaker pronunciation model appears to be a more viable target, especially when training future English language teachers, interpreters, etc. But which native-speaker model should we choose? And will that be what the learners themselves want?

The paper reports the results from several surveys conducted by us over the last few years among tertiary-level students of English at Sofia University. The students’ opinions and preferences regarding a number of pronunciation issues are compared with the findings of similar studies surveying the opinions of learners of English in other European countries (Poland, Spain, etc.). We comment on the steady tendency for Bulgarian university students of English to favour a standard British (English) English pronunciation model – RP – as their target.
1. Introduction

At the beginning of the 21st century English has undoubtedly established itself as the language of international communication. This global spread has taken place with unprecedented rapidity because more and more people need to use English for social, educational, and professional reasons in all kinds of contexts, locally and internationally and because the development of electronic technology has effectively diminished the constraints of space.

Regarding users of English around the world the basic distinction is between native speakers and non-native speakers of English, that is, those born to the language and those who have learnt it through education. Thus speakers of English in the world form three distinct groups: ENL (those who speak English as a native language), ESL (those who speak English as a second language), and EFL (those who speak English as a foreign language). It is often observed that second and foreign language speakers outnumber those for whom it is the mother tongue, which means that millions of interactions take place in non-native English where non-native speakers talk to other non-native speakers. Therefore, Jenkins (2000) speaks of a fourth group of speakers of English – EIL - those who speak the so-called international language. She thinks that the term “EFL” is appropriate only to describe interactions between native and non-native speakers of the language, and that the term EIL should be applied to situations in which non-native speakers communicate with other non-native speakers. Alternatively, one can refer to the latter group of speakers by reversing the second and third letters of “EFL”, to arrive at ELF, or “English as a Lingua Franca” (Jenkins 2000:10-11).

The growing varieties of English and its expansion in the world have generated a lively discussion of the issue of the choice of English accent to be taught to foreign learners, whether one should use a native speaker model or a non-native one. The decision on what model(s) of English pronunciation to teach students is a complex one because of the great variety of accents. First, because of the social and regional varieties of English in the British Isles people from different social classes and regions have different pronunciation. Second, English is spoken all over the world as a second or foreign language, or as a lingua franca for international communication. The pronunciation of all these people is influenced to a different extent by their first language, as a result of which their pronunciation is particularly variable. Based on these differences in pronunciation we speak of “Indian”, “Malaysian”, “Bulgarian”, “Japanese” English, etc.
That leads us to the point that it is important “to distinguish between a model as a ‘target’ and as a ‘point of reference’” (Hewings 2004: 13). We speak of the model as a target when we have some standard of pronunciation, such as GA or RP, to which the students aspire or which we have selected as a goal for students. A model is a point of reference when it is recommended as a guide to English pronunciation with the understanding that there might be variation from this model as long as it does not interfere with effective communication.

Perhaps the most important outcome of recognizing the complexity in the setting of goals in teaching pronunciation model(s) is to find out what our students know about the different varieties of pronunciation, and to encourage them to consider who they want to sound like when they speak English and what pronunciation model(s) they prefer, as well as what level of proficiency they would like to achieve in their English pronunciation.

2. Background

Although teaching English pronunciation, be it in the form of the old “Remedial Phonetics” classes or, much more often, in some more up-to-date form, constitutes an inextricable part of the academic curriculum, one could argue that at least to some extent this may be due to tradition, and that teaching tertiary-level students native-speaker accents should be abandoned. The latter idea has recently found a number of proponents as well as critics (see for example Jenkins 2000, 2004, Seidlhofer 2001, and Wells 2005, among many others).

For many years Received Pronunciation (RP) and General American (GA) have been the two models traditionally adopted in teaching pronunciation to foreign learners. They have been widely used because of some obvious advantages they have: they are easy to learn and describe for learners, understood by all native speakers, pedagogically tested, codified for teaching purposes (especially RP) and recognized as the English of the media (BBC and CNN). They constitute the basis for the big majority of teaching materials, recordings and available pronunciation dictionaries.

However, a number of scholars have made claims that “It is now generally accepted that the target of a ‘native-speaker’ pronunciation is unachievable for the majority of learners of a second or foreign language, even if a native speaker variety is the target model chosen.” and “that it is neither necessary nor even desirable” (Hewings 2004: 13). According to this view, learners of English should aim to achieve a pronunciation that is easily understandable in most situations by both native and non-native speakers. These scholars also believe that “It is also important to
remember that a person’s pronunciation (of both their first and other languages) contributes significantly to the impression of their identity that is conveyed to others” and argue that some students want “to retain identifiable traces of their national or first language identity” (Hewings 2004: 14) when speaking English. They find that a more appropriate and reasonable goal, for most learners, is to achieve an English pronunciation that is understandable in international communication, ‘but retains unobtrusive features of a non-English accent.’

One of the new proposals is to teach foreign students a simplified version of the language, i.e. to establish a necessary core which students will find easier to learn than the two standard models of pronunciation. Jenkins puts forward the Lingua Franca Core (LFC) of ELF to replace the standard models taught to the students so far (RP and GA). She argues that learners of English do not need to acquire a native accent because they mainly communicate with non-native speakers.

There were similar proposals in the past, such as Gimson’s RIP (Rudimentary International Pronunciation), or Crystal’s WSSE (World Standard Spoken English). The underlying assumption of LFC is that the standard accents contain features that learners do not need for the purposes of international communication. The acceptance of a “phonological core” would make learning important elements of the language easier, would help non-native speakers better understand each other and would enable learners to preserve their local identity.

The pros and cons of teaching a “simplified form of international English pronunciation” have been widely debated since the publication of Jenkins’s book. On the face of it, the LFC allows us to save a lot of classroom time and effort, and yet make certain that our students will be able to understand and to be understood, at least when they speak to other non-native speakers. But, as aptly pointed out by Wells (2005), students should not be made to choose between EFL and ELF. Both at tertiary level as well as later at work, young people should be able to use English both for communication with other non-native speakers as well as with native speakers of the language.

Another argument to take into consideration is that non-native learners are not only speakers who might benefit from making their pronunciation goals as simple as possible, but also listeners. Peter Trudgill claims that “Smaller phonological systems may make things easier for the non-native learner as producer: there is less learning to carry out and less articulatory effort involved. However, they also make life much more difficult for the non-native learner as listener by reducing the amount of phonetic information available for processing.” (Trudgill 2005: 222) Inability to
produce a phonemic contrast lessens the chances of hearing it also. Teaching non-native learners to produce contrasts will help their comprehension because their ability to perceive phonemic distinctions improves the recognition of words (Trudgill 2005).

In her paper “Brave New English?” (2001), B. Seidlhofer observes that in many English Departments throughout Europe students learn in their applied linguistics courses about EIL and are urged to question native speakers’ ownership of English. “But then, these very same students go from their applied linguistics course to a room next door for their English language class, where they are taught, and are usually eager to learn, English idiomatic expressions and proverbs originating in the UK and the US as well as American English and English English intonation patterns, th-sounds and flapped t’s; …” (Seidlhofer 2001: 43). Maybe a key expression in this quote is “…are usually eager to learn…”!

Prof. Seidlhofer goes on to admit that “…there may be learning purposes for which adhering to native-speaker models is a valid, or at least arguable, option.” (ibid.), and university English departments are likely to be one of the first among these. If students are indeed “eager to learn” native-speaker pronunciation, then is it right to deny them the opportunity to do so? Given that even the most ardent proponents of ELF acknowledge the fact that it is ultimately up to the learners to choose their goals, we must accept that “even a learner whose target community is an ELF one may prefer a native rather than an ELF variety as their goal” (Jenkins 2004: 65).

Also, the Lingua Franca Core does not take into consideration the typological differences among languages: what is simple for Bulgarians might not be simple for German, or Turkish, or Chinese learners of English. This typological perspective makes arguable the existence of one common core for learners speaking different first languages.

In this paper we argue against teaching a form of “International pronunciation” to university students of English, and report results from several surveys conducted by us over the last few years among tertiary-level students whose major is English and American Studies at Sofia University. The students’ opinions and preferences regarding a number of pronunciation issues are also compared with the findings of similar studies in other countries.
3. The investigation

3.1. Questionnaire

In October 2009 we surveyed the knowledge about and attitudes to English pronunciation of freshers at the Department of English and American Studies, University of Sofia. The questionnaire was compiled by drawing on our previous experience and on similar research reported in the literature (Celce-Murcia et al. 1996, Pennington 1996) and adapted in accordance with the aims of the research. The Questionnaire can be seen in the Appendix.

3.2. Subjects and procedure

Ninety first-year students were asked at the start of their first Practical English Phonetics class to respond in writing to the eight questions in the questionnaire. The students’ age was between 18 and 31, the majority of them being between 18 and 20. Twenty-three students were male and sixty-seven were female.

3.3. Results and discussion

The aim of Questions 1 and 2 was to collect information about the students’ experience in learning English. The responses to Question 1 show that they had learned English for an average of 9.5 years before coming to university. During this period, they were taught predominantly by Bulgarians, and only occasionally by native (British or American) teachers. A few students reported that they had communicated with friends or colleagues from Britain, the USA, Australia, or that they had worked abroad for a couple of months. Two students, however, said that they had had virtually no contact with native speakers of English prior to university.

Question 2 asked students about the main accent to which they had been exposed. Figure 1 sums up the replies. An equal percentage of students (25%) reported having had experience predominantly with British English and with American English, respectively. However, the majority of students (40%) described their previous experience as mixed exposure to both accents. Some students from this latter group commented that they had used British English at school, while outside school they had been exposed mostly to American English – through movies, TV, other forms of popular culture, the Internet, etc.
Some students also reported having had experience with Australian English (2 students), Irish and Scottish English (5 students each), South African, Canadian, Indian and Italian English. These answers have not been included in Figure 1.

**Question 3** wanted to find out whether the students had had any pronunciation instruction in primary or secondary school. The results are shown in Figure 2.

Twenty-two students (24.5%) replied that their teachers had indeed taught them about pronunciation, including one student who said that this
had been done with the help of multimedia. Twenty-seven students (30%) reported that they had been taught only the basics, while forty-one students (an overwhelming 45.5%) said that at school they had received no (explicit) pronunciation instruction.

**Figure 3.** Communication problems attributable to pronunciation.

Question 4 asked the students whether they had ever had communication problems due to pronunciation. Their answers are presented in Figure 3.

While a majority of 57 students (approximately 63%) answered that they had never experienced any communication problems, another 22 (approximately 25%) admitted to having been misunderstood while working abroad, talking to native speakers abroad, because of mispronouncing a particular word, etc. These results, in our opinion, emphasise the need to train tertiary-level students to be able to communicate efficiently with all sorts of speakers, including native English speakers. Therefore the idea of teaching our students the Lingua Franca Core rather than encouraging them to aim at native speaker pronunciation models does not present an acceptable alternative.

The next three questions aimed at investigating the need to teach specific aspects of pronunciation. We have commented on the responses elsewhere (Dimitrova and Chernogorova 2010), therefore we will not discuss them in any detail here. Instead, we will focus on Question 8, which was designed to collect data about the students’ accent preferences. It read as follows:
8. In your opinion, which groups of speakers or well-known individuals have
   (a) the best / worst pronunciation of English?
   (b) the type of English pronunciation that is the easiest / hardest to
       understand?
   (c) the type of English pronunciation that is the easiest / hardest to
       imitate or learn to speak oneself?

The question was borrowed from Pennington (1996: 30) and subsequently re-phrased. However, some students answered only the first part of each sub-section; therefore, when calculating the percentages, we have taken into account only the actual number of replies received. Thus, in the results reported below, 100% represents the total number of answers received for that particular part of the question, but is not always equal to 90 (the overall number of students who took part in the survey).

At the time when they participated in the investigation, many of the students were unaware of terms such as RP or GA; therefore, when discussing their answers, we will use the term “British English” as a cover term, meaning “RP” or “non-regional English pronunciation”, “BBC English”, etc., and “American English” - as a cover term meaning “GA”, “Network American”, etc.

**Question 8a** asked the respondents what they thought was the best and the worst pronunciation of English. The answers should of course be interpreted with caution, since the question involved such impressionistic labels as “good/best” and “bad/worst” whose meaning in this context is not unambiguous. Indeed, one student aptly commented that “There is no such thing as bad pronunciation”! Still it is quite interesting to note that, as far as “the best English pronunciation” is concerned, out of the 81 answers which were given, only 3 (3.5%) pointed to non-native accents as being “the best”, whereas an overwhelming majority (57 students, or 70%) thought that one or another of the native accents of English was “the best”. Figure 4 summarises the results; the column “other (native)” stands for Australian, Scottish, and Irish English; “non-native” stands for Scandinavian, Bulgarians, and Mid-Atlantic English.

The answers about “the worst” English pronunciation were more diverse and hence more difficult to summarise: British English was mentioned only once, American English – 8 times, Australian English – twice, Scottish English – 7 times, Irish English – 5 times, alongside “the speech of cowboys”, Cockney, “English spoken in the Bahamas”, the West Indies, African-American English, etc. Thirty-two respondents answered by naming native accents, but three-fourths of these were
different than standard British or American English. Another eighteen students named non-native accents (Bulgarian English, Russian English, Italian English, French English, German English, English spoken by Arabs, Latin Americans, etc.), and seven students said they did not know.

**Figure 4.** Answers to question 8a.

![Bar graph showing percentage preferences for different English accents.](image)

**Figure 5.** Answers to question 8b - the English pronunciation which is the easiest to understand.

![Bar graph showing percentage preferences for different English pronunciations.](image)
Question 8b wanted to find out which accent(s) Bulgarian students found easy (respectively, difficult) to comprehend. All 90 students answered this question. The easiest accent was American English (49 students, or 54%), followed by BrE (8 students, or 20%). “Native English”, Australian and Cockney were each mentioned once, “the pronunciation of foreigners” – by 3 of the students; together with one student who found Bulgarian English to be the easiest accent to understand, only 4.5% of our respondents seemed to think that foreign-accented English was easier to comprehend than a native accent. (see Figure 5).

These results are similar to our own previous findings (Dimitrova and Filipov 2005, 2007): in a study in which tertiary-level Bulgarian students evaluated several accents in terms of a number of voice and personality traits, GA was given the highest score for “comprehensibility”. The scores given to GA in that study were higher even than those for Bulgarian English, which received the second highest scores.

Together, these findings seem to cast doubt on the claim that “International English” is “easier” for foreign learners: it is not, at least as far as tertiary-level students of English in Bulgaria are concerned.

**Figure 6.** Answers to question 8b - the English pronunciation which is the hardest to understand.
The second part of question 8b about the English pronunciation which is “the hardest to comprehend” is again characterized by a diversity of answers, but typically Scottish English (25 answers, or 30%) and Irish English (17 answers, or 21%) are the accents which our students find difficult to understand. British English, however, is also mentioned – by 11 students, or 13%, along with American English – 6 students, or 7%. Other native-speaker accents which students find hard included Australian English, New Zealand English, Welsh English, etc. Non-native accents, on the other hand, were mentioned by only 3 students. (See Figure 6.)

It should be pointed out, however, that these latter findings should be interpreted with a certain amount of caution, because the formulation of the question may have suggested to the students that they should think about native rather than non-native English accents (although the fact that non-native pronunciation was in fact mentioned suggests that this is not very likely).

Finally, question 8c investigated students’ opinions as to the relative ease/difficulty involved in mastering an English accent. While only 61 students (about two-thirds) answered this question, 59% of them (36 students) replied that they thought that American English was the easiest to learn to speak. British English, on the other hand, was the choice of 20% (12 students). Figure 7 sums up these results.

**Figure 7.** The easiest accent to learn to speak
The second part of question 8c asked students about the English accent which they thought was the most difficult to learn to speak. The “winner” here was British English, which was named by about a third of the students (23 out of 71 who answered the question, or 32%), followed by Scottish English (27%), Irish English (15.5%), etc. American English, on the other hand, received a mere 1.5% (only one student thought it was a hard accent to learn to speak).

**Figure 8.** The hardest accent to learn to speak

4. Conclusion

The answers given by the tertiary-level Bulgarian students of English who took part in the survey reveal a number of interesting tendencies.

4.1 Native accents are clearly preferred by the students to non-native accents: it is native accents that are perceived as being “the best”. It should again be pointed out, however, that the native vs. non-native distinction was not stated or implied in any of the questions, so students may have assumed that it was only their opinions regarding native accents that were being investigated. On the other hand, a number of non-native accents were mentioned, and described as “bad”. Among the native-speaker accents, “British English” was viewed as “the best” accent of English.

4.2 American English is the accent which students think is the easiest to understand. It is also the accent which they find most easy to imitate, or
to learn to speak themselves. British English, on the other hand, is in students’ opinions the hardest accent of English to acquire.

**Figure 9.** Accent preferences: 47 first-year Bulgarian students of English (October 2009)

![Bar chart showing accent preferences](chart)

4.3 A few weeks after the investigation, we conducted a short follow-up survey: we asked a random selection of the same students to name the accent they wanted to model their own pronunciation on. Forty-seven students participated in the follow-up investigation. They had to complete the sentence “I would like to speak English like...”. About two-thirds of them (57.5%) said that they wanted to sound British, another 15% replied that they wanted to speak like an American, 2 students (4%) wanted to sound Australian, while about a quarter (11 students, or 23.5%) replied that they hadn’t decided yet. These answers are summarized in Figure 9.

These findings are in line with results from previous surveys which we have conducted over the last few years. Dimitrova and Filipov (2007) reported that Bulgarian students of English evaluated GA as the easiest to comprehend, yet when they were asked what accent they hoped to acquire by the end of their English BA studies, about two-thirds of them (64.9% of the first-year students and 60.3% of the second-year students who took part in that earlier survey) answered that they wanted to speak with an RP/BBC/Standard British accent: a really interesting result, considering that the same students also evaluated GA as being more “pleasant”, “attractive”, as well as more “comprehensible” than RP, Scottish, New Zealand, and Bulgarian English. Dimitrova (2008) also reported that, in a survey conducted at the end of the first year of study, a group of Bulgarian
students of English said that they preferred RP (59%) to GA (23.5%). When asked to justify their preferences, these students replied that RP sounded “beautiful”, “classier”, “educated”, and “sophisticated”, whereas GA was described as “familiar” and “easy to understand”. A couple of students mentioned explicitly that their choice of accent was connected with future plans to study or work in Britain or the US.

A brief comparison with similar research carried out elsewhere shows that our results are not surprising or unique. Studies such as those presented at the Poznan Linguistic Meetings (PLM 2003 and 2004) and published in Dziubalska-Kolaczyk and Przedlacka (2005) seem to paint a similar picture. Most of these studies, like our own investigations, focus on university students’ pronunciation goals, opinions, attitudes and preferences. What the big majority of these students seem to want is to be taught native-speaker models; of these, standard British models seem to be preferred not only by Bulgarian but also by Polish learners (Waniek-Klimczak & Klimeczak 2005, Janicka, Kul & Weckwerth 2005), although in the first of these studies students also found RP easier to understand, which was not the case in the study reported here. Cenoz & Lecumberri (1999) also demonstrate that RP is easier than GA for Spanish and Basque learners of English. Research reviewed by Remiszewski (2005: 297-303) reports preferences for native-speaker accents expressed by students from Korea, Japan, etc.- though not by students from South Africa, Pakistan and India, where almost exactly the reverse tendencies are observed.

The results of Janicka et al (2005) are much more similar to our own findings: they report that Polish students at the School of English, Adam Mickiewicz University in Poznan considered AmE to be “easier to learn and master” than BrE; at the same time some students said that they “wanted to confront the challenge”, and hence chose to acquire a BrE pronunciation (Janicka et al 2005: 255).

Such findings challenge one of the major reasons traditionally given for choosing RP as a model for the pronunciation instruction of foreign learners of English, namely, the ease with which it is understood. Notwithstanding a long tradition of teaching RP to Bulgarian learners of English, and in spite of the students’ claims that they have so far been equally exposed to British and American English, they clearly find British Received Pronunciation a relatively difficult accent to comprehend. Yet, in spite of that, they appear to prefer it as a model to adhere to. The motivation behind this choice remains to be investigated. But irrespective of the students’ reasons for choosing RP – or, for that matter, GA – it seems that the tradition of teaching native-speaker pronunciation models to tertiary-level students of English in Bulgaria, and in many other parts of
the world, is likely to continue unabated at least for the time being because this is what many instructors as well as the overwhelming majority of their students want.

References


Appendix 1 - Questionnaire

1. Please describe your experience in learning English (e.g., how long, where, what kind of teachers, exposure to native speakers of English in your country and abroad, etc.)

2. Have you had more experience with British or with American English (or some other variety, such as Australian, Irish, Scottish, etc.)? Please explain.

3. Have any of your previous English teachers taught you about pronunciation? If yes, please give some details.

4. Have you ever been misunderstood because of your pronunciation? If yes, please give some details.

5. Do you think your pronunciation needs improving? If yes, please specify what aspects you would like to improve (e.g., individual sounds, stress, rhythm, intonation, other aspects).

6. Do you ever use the pronunciation key or guide in your dictionary to get an idea of how a word is pronounced?

7. Are you familiar with a phonemic alphabet or any phonetic/phonemic symbols? If yes, please explain and/or give examples of such symbols.

8. In your opinion, which groups of speakers or well-known individuals have
   (a) the best / worst pronunciation of English?
   (b) the type of English pronunciation that is the easiest / hardest to understand?
   (c) the type of English pronunciation that is the easiest / hardest to imitate or learn to speak oneself?
ENGLISH PRONUNCIATION NORMS
AND THE CASE OF RUSSIAN ENGLISH

GALINA M. VISHNEVSKAYA

Outline

“Variability” or “variation” of speech (pronunciation in particular) is determined by a number of intralinguistic and extralinguistic factors. In the teaching of English to non-native speakers the choice of the pronunciation norm in view of the enormous variety of the pronunciation types in the target language may be quite a dilemma for a teacher and a student. Now that English has become “a global language” and by this standing as “a lingua franca” has given rise to a number of “Englishes” (Crystal 1997), there are two conversely opposed and, at the same time, inseparable trends in the English language development of to-day: the ‘pull’ towards the standard pattern and the ‘push’ towards the modes of language variation. The quality of speech supplied by the label “foreign accent” can be regarded as one of the additional speech variation factors determined by the influence of the mother tongue and revealed through the deviation from the "pattern" which is, in our case, the pronunciation norm of English. Our chief concern in the present research is the study of prosodic variables of accented speech as opposed to standard native speech suprasegmental characteristics. The data has been collected from both natural settings and controlled experiments.

1. Introduction

The pronunciation norm of English itself constantly changes under the influence of numerous speech variation factors (Wells 2006). However, it remains to be an ideal abstraction pattern which serves as an indispensable model in the teaching pronunciation practice of classroom (artificial) bilingualism. With the variety of English accents existing now there, naturally, arises a question: Which accent to choose for teaching and learning purposes?
Jennifer Jenkins (1998) was one of the first scholars who insisted that the acquisition of a native-like accent is no longer the ultimate objective of the majority of learners. And the primary motivation for learning English in the 21st century of globalization has stopped to be communication with native speakers alone, but communication with non-native speakers of English using it for purposes of successful, mainly professional, communication in many spheres of life.

According to our poll (2010) most of the students at Ivanovo State University, Russia, would choose British pronunciation standard as their learning model. School teachers of English still prefer British English as a teaching norm for the reasons of clearness of enunciation and better understanding.

The pronunciation norm is constantly changing under the influence of the speech variation factors but it remains to be an ideal abstraction pattern, which serves as an indispensable model in the teaching pronunciation practice. The most significant problem arising from the situation of artificial bilingualism is the right choice of the pronunciation standard of English. The pronunciation standard should be easily understood and positively accepted in the prime spheres of social life, and, finally yet importantly, well supplied with quality teaching and learning materials. This proviso will surely enhance the chances of a non-native speaker to be rightly understood and adequately treated in the English speaking community. Intonation performs most important functions in human communication: expressive, emotional, logical, esthetic, etc. It is said to be the ‘soul’ of the language, while the pronunciation of the sounds is its ‘body’ (Kingdon 1958). It is a well-known fact that English speakers are able to make a good deal of allowance for imperfect sounds, but they are much less able to make the same allowance for mistakenly used intonation.

2. Theoretical background

For any spoken language, there are prominent regional, social, and individual differences in the way the language is pronounced by different people. Each distinguishable type of pronunciation is called accent. Accent is “…a particular way in which a group of people collectively pronounce a language” (Trask 2007: 3). It is the “…cumulative auditory effect of those features of pronunciation which identify where a person is from, regionally or socially” (Crystal 1995: 2). Practically everybody has an accent. Moreover, every one of us regards accents differently: some accents are familiar, some may sound startling, and others are thought as prestigious.
In Britain, “...the single most prestigious accent is Received Pronunciation, or RP, an accent which seems to have arisen in the prestigious ‘public schools’ (private schools) in the nineteenth century, and was adopted as the ‘voice of the BBC’ in the 1920s. This accent is not associated with any particular region, though it is structurally most similar to certain accents of the south-east coast of England. No more than 3 per cent of Britons speak with an RP accent, though many more have a near-RP accent which differs only in a few particulars” (Trask 2007: 3).

Estuary English serves as an example of an English accent. It is a name given to “...a wide spectrum of accents used in southeastern England and falling somewhere between broad Cockney at one extreme and unmistakable Received Pronunciation at the other. Associated particularly with upwardly mobile working-class speakers eager to identify themselves with ‘popular culture’. Some observers have suggested that Estuary English may be slowly replacing RP as the prestige accent of southern England” (Trask 1996: 134). RP has several labels: D. Jones defined it as ‘public school pronunciation’; sometimes it is called ‘a BBC accent’, or ‘an Oxford accent’; those who dislike it “...may call it a plummy, lah-di-dah, Kensington or cut-glass accent” (Trask 1996: 302). Several varieties of RP are often distinguished: A. C. Gimson (1989) recognizes conservative, general, and advanced varieties of RP; J. Wells (1982) distinguishes mainstream RP, U-RP, and adoptive RP. There is also Estuary English, mentioned above, and near-RP (Wells 1982). RP is the accent usually taught to foreign learners of English.

Language contacts and the situation of massive bilingualism (and polilingualism) introduced new varieties of English. The English language as a lingua franca is modeling now a speech situation where an individual or community controls two (or more) languages. There are many types of bilingualism (Beardsmore 1986): it may be natural vs. artificial, full vs. partial, simultaneous vs. sequential, compound vs. co-ordinate, early vs. late, etc. Both co-ordinated bilingualism and compound bilingualism develop in early childhood and are classified as forms of early bilingualism. Late bilingualism develops when a second language is learned after age 12. Late bilingualism is defined in contrast to early bilingualism, because the former is developed after the critical period for language learning. The case of late bilingualism is typical for the situation of classroom bilingual education.

Bilingualism is closely connected with the linguistic phenomenon known as ‘language interference’. It is the process whereby a speaker introduces errors into one language as a result of contact with another language, typically while learning a foreign language or while living
within a multilingual situation; it is also called negative transfer. In his article on overcoming phonetic interference, John Wells writes (2000: 27): “When we encounter a foreign language, our natural tendency is to hear in terms of the sounds of our own language. We actually perceive it rather differently from the way native speakers do. Equally, when we speak a foreign language we tend to attempt to do so using the familiar sounds and sound patterns of our native tongue. We make it sound, objectively, rather differently from how it sounds when spoken by native speakers”.

Because of the language contact between English and Russian in the speech of a bilingual person (in our case - a Russian learner of English), there appear certain speech defects obvious to any native speaker. The language teachers call them ‘mistakes’. The impaired pronunciation effect of non-native speech is called ‘foreign accent’. L. V. Sheferba spoke of the two types of pronunciation mistakes in the speech of foreign language learners (1953): phonetic (‘oshibki vygovora’) and phonological mistakes (‘zvukosmysloviye oshibki’). Phonological mistakes are the ones that result into the change of meaning of a pronounced word (e.g. ‘pen’ instead of ‘men’). Phonetic mistakes are those that betray the origin of the non-native speaker, displaying the unusual quality of segmental and suprasegmental characteristics of speech. Of all the suprasegmental characteristics of oral speech intonation is most obviously betraying the speaker’s origin, social status, educational background, individual peculiarities of pronunciation. Intonation is the most communicatively charged component of suprasegmentics. Since intonation is capable of transmitting a whole range of linguistic and extralinguistic meanings the study of it is of primary importance for the practice of teaching and learning foreign languages.

The first serious interest in the study of intonation was observed in the 90-ies of the XIXth century when European scholars started to recognise its great role in grammar, stylistics, poetics, logic, psychology, etc. There emerged an opposition between “Ohrenphilologie” and “Augenphonologie” (Sievers 1912). Oral speech was opposed to written text for the first time and it gave rise later to many intonation studies. Many Russian scholars of the past – V. N. Vsevolodski-Gerngross (1922), G. P. Torsuyev (1939), N. I. Zhinkin (1946), A. M. Peshkovski (1956), V. A. Artemov (1956), etc. – were keen upon the study of intonation. N. I. Zhinkin (1946), for instance, performed a research on psychological and pedagogical aspects of intonation in connection with the study of the general problems of speech expression. Later, in 1958, the scholar wrote his famous work “Mechanisms of Speech” where his views on intonation were clearly pronounced. He considered intonation to be an integral part of any speech
act. And he understood intonation as a complex of various phonetic features. The Russian linguist G. P. Torsuyev (1939) carried out the first experimental work on Russian intonation as compared to English. This work laid foundations for further extensive and intensive intonation research in the country which resulted into the creation of the theory of intonation based on the experimental investigation of Russian, English, German, French and other languages.

intonation studies in Great Britain from the very start were to a great extent pedagogically oriented. British phoneticians have contributed a lot into the teaching of English intonation to ‘overseas students’. The most well-known book of this kind appeared at the beginning of the previous century: it was Harold Palmer’s “English Intonation with Systematic Exercises” (1922). Later more intonation text-books of distinction appeared (Armstrong, Ward 1926; Kingdon 1958; Halliday 1967; O’Connor, Arnold 1961, 1973, etc.). The work by David Crystal “Prosodic Systems and Intonation in English” (1969) has turned out to be the most fundamental work on the theory of suprasegmentals including the description of English intonation (British English). At the end of the 20th century intonation has stopped to be the ‘Cinderella’ of the linguistic sciences. The year of 1986 has turned out to be the Vintage Year in the intonation research: there appeared three significant books — a two-volume bible’ by the American linguist Dwight Bolinger (“Intonation and its Parts” and “Intonation and its Uses”), a monograph by the British scholar Alan Cruttenden (“Intonation”) and a book by the German phonetician Elizabeth Couper-Kühlen (“English Prosody”). The most recent works on English intonation (Chun 2002; Wells 2006, etc.) point to a renewed interest in this department of phonetic sciences.

The majority of Russian phoneticians view intonation as a complex unity of several components, formed by communicatively relevant variations in (Vassilyev 1970: 290): (1) voice pitch, or speech melody; (2) the prominence of words, or their accent; (3) the tempo (rate), rhythm and pausation of the utterance, and (4) voice timbre.

Intonation mistakes are the ones that are most difficult to overcome for a foreign language learner since intonation in his/her native tongue is the primary feature of oral speech being acquired by a child in the very early stages of his/her life, before he/she actually starts using words. It is rooted in a child constituting the basis for the further development of his/her oral speech, both in its phonetic form and the meanings conveyed by it. The ‘pre-word’ period of a child’s life is full of intonation impressions coming from his/her mother and the people surrounding the child. The child reactions to grown-ups’ speech are mainly represented by vocalizations
supplied with voice inflections imitating the intonation patterns heard by the child. Any foreign language learner is deprived of this long ‘pre-word’ period of language acquisition and comes to learning the intonation of the target language much later than learning the words of the language. Besides, the language learner is burdened with his native tongue intonation knowledge lying dormant in him and influencing him in the perception of the foreign language intonation patterns. The latter ones are conveying linguistic meanings different from the native tongue, though sometimes seeming almost the same in form. S. I. Bernstein wrote about ‘patterning’ the oral impressions of a foreign tongue in the process of its learning (1975: 7). The foreign language learner is often misled by the fact that there are universals in intonation the melodical component of which is based on the fall and rise of the voice pitch practically in every one of the world languages. The language learner does not at first feel the difference in the tone configuration and its other parameters pertaining to the suprasegmental characteristics of the target language.

On the level of performance, the foreign language learner is also under the strong influence of his native tongue intonation habits, which leads to the wrong suprasegmental organization of the utterance in the choice of melody, peaks of prominence placement, placement of pauses, rhythm arrangement of an utterance, etc. The distorted perception of foreign speech determined by the ‘phonological sieve’ (the term was originally used by N. S. Trubetzkoy) of the native tongue can explain the causes of pronunciation mistakes resulting into the foreign quality of oral speech (1960: 60–62).

The phenomenon of accented speech has long been of interest to scholars from a variety of orientations. Applied linguists, for instance, need to account for the fact why foreign accents occur as they develop theories of second language acquisition. Researchers specializing in speech perception and production seek to identify the specific phonetic properties of accented speech and to explain those properties using theoretical models that give us insights into the way speech is represented psycholinguistically. In addition, pedagogical specialists are concerned with methods for remedying aspects of an accent that impede effective communication. Some researchers are also interested in the problem of negative social evaluation due to discriminatory attitudes towards accents in the society. For these reasons the study of foreign accents - their causes, characteristics, and impact on communication - is an important area of inquiry within the field of linguistics.

Any language is a dynamic category, constantly changing under the influence of several variation factors: territorial, social, cultural, individual,
etc. Language contact impact is one of the strongest factors of language variation. The quality of speech supplied by the label “foreign accent” is determined by the influence of the mother tongue; it is revealed through the deviation from the pronunciation «pattern», easily detected by the native speaker’s ear (Vishnevskaya 1993). Intonation is, perhaps, most obviously betraying the non-native origin of the speaker, its lingering effect ever staying with the bilingual speaker.

3. Research results and discussion

3.1. Evaluation of foreign accent

In our research, we tried to discover the effect produced by foreign accent upon native speakers (American respondents) and upon non-native speakers (Russian respondents). The study concerned adults’ perceptions of other people’s accents. Interviews were conducted with 40 American University students (University of Boston, USA) and 50 Russian University students (Ivanovo State University, Russia).

Questionnaire 1 for English speaking students consisted of the following questions:

- Are you a native speaker of English?
- Is it easy for you to identify the different American regional accents?
- Are you able to understand the words spoken by someone with an American regional accent?
- Do you like listening to someone who speaks with an accent?
- Can you identify a speaker’s native language when listening to the person speaking English?
- Are there any accents which sound pleasant or unpleasant to you?
- Can you identify the speech of the foreigner as having a ‘heavy’ accent?
- Can you identify the speech of the foreigner as having a ‘slight’ accent?
- Do you become frustrated when speaking with someone whose English is slightly accented?
- Do you become frustrated when speaking with someone whose English is heavily accented?
- Why do you become frustrated?
- Do you become frustrated when listening to a radio or TV broadcaster who speaks with a foreign accent?
Would you prefer a foreigner to speak with an accent?
If you spoke a foreign language, would you want to speak without an American accent?

The study of the evaluative reactions of native speakers towards non-native speech is important because it deals with the problem of effective human communication. Undoubtedly, the speech behaviour of a foreigner affects a native speaker’s attitudes, based on liking or disliking accented speech and certain types of it. The American respondents all admitted their ability of detecting accented speech and identifying some of the accents. Both regional and foreign accents of English can have either positive or negative evaluative reactions on the part of the English native speakers. The speakers’ productions were almost always judged to be accented. As native speakers themselves admitted in the course of our investigations, some regional accents of English can be overwhelming; they stigmatize people as having “nasty accents”. However, in general, American respondents have displayed a great degree of tolerance towards accented speech. Most of them said that they would not like the speakers to get rid of their accents. As far as their own American accent is concerned, none of the respondents wanted to lose it in speaking a foreign tongue.

Questionnaire 2 for Russian speaking students, learning English as a foreign language, included the following questions:
Can you easily detect traces of foreign accent in somebody’s speech?
Do you like the sound of accented speech?
Do you ever think of accents as ‘pleasant’ or ‘unpleasant’?
Can you detect the degree of accent in somebody’s speech?
Can the degree of accent influence your opinion of the accent as being ‘pleasant’ or ‘unpleasant’?
Does accented speech hinder your communication with a non-native speaker?
Do you like hearing accented speech on the radio or television?
Would you like a foreigner to speak your language without any accent?
Would you like to speak the foreign tongue you are learning without a trace of your native accent?

Russian learners of English in their answers to the questions differed from the American respondents in that they all wanted to get rid of their Russian accent in English. Their attitudes towards foreign accent in Russian speech were more often, than in the group of American
respondents, negative: accented speech was rather often defined as ‘funny’ or ‘unpleasant’. Interestingly, all the Russian respondents expressed their ardent desire to speak ‘accentless’ English. However, most accents, according to the responses of the students in both the groups (native and non-native speakers of English) were characterized as ‘charming’, ‘interesting’, ‘distinctive’, ‘attractive’, ‘amusing’. Some accents were labeled negatively as ‘irritating’, ‘distracting’, ‘boring’.

In our pilot study, the respondents displayed a definitely negative reaction to the heavily accented speech. A slight accent, on the contrary, evoked positive attitudes. These observations led us to distinguishing between two types of ‘mistakes’ in accented speech: phonetically and communicatively relevant. The former betray the language learners non-native origin (foreign-sounding); the latter genuinely affect the learners’ ability to communicate.

3.2. The case of Russian accent

The analysis of the pronunciation error variables in English speech of Russian learners is not only a matter of the phonetic specifics of Russian English. The communicative effect of foreign accent as a non-native quality of oral speech is crucially important in certain social contexts. The communicative effect of an individual’s accent upon a native speaker, as many researchers have shown, can be either positive or negative depending on the degree of the accent, social attitudes to it, individual preferences (Honey 1989; Вишневская 1993; Lippi-Green 1997; Jilka 2000; Derwing, Munro, Rossiter 2002; Huckvale 2007, etc.). Of all the suprasegmental (prosodic) characteristics of oral speech intonation is known to be of primary importance communicatively. That is why the analysis of accented speech on the intonational level is one of the central issues in foreign accent research.

Segmental characteristics of Russian accent are most obvious at the early stages of language learning, with mistakes of both phonetic and phonological character in vowel and consonant production. Suprasegmental characteristics of Russian English (in syllable formation, word accent, intonation) are more pronounced at later stages of language learning and are more difficult to overcome. There is a definite pattern of foreign accent in English intonation marking the overall melodical contour of an utterance, peaks of prominence distribution, rhythmical arrangement, temporal variables, quantitative and qualitative character of pausation and timbre speech characteristics. Foreign accent is marked by certain general
deviations from the norm revealing, on the one hand, the non-native character of speech, and on the other, specific prosodic features pointing at the origin of the non-native speaker.

Our description of intonation variables in Russian English (in the speech of Russian bilinguals) is based on the results of auditory analysis (phonetic experts’ opinion). Different functional styles of oral English were analyzed: lecture style, reading style, conversational style. It has been discovered that Russian bilinguals (students of senior university courses) tend to have errors both in intonation perception and intonation performance.

Errors of intonation perception were characterized by detecting a smaller number of pauses, misplacing intonation breaks, using wrong distribution of stress elements, marking mistakenly nuclear tones (especially complex and compound tones), giving obscure picture of prenuclear melodical patterns, etc. Intonation performance revealed two main groups of mistakes detected in the speech of Russian learners of English: qualitative and quantitative. The distribution of intonation patterns in the speech flow showed a tendency to simplification and revealed ‘hidden’ mistakes of a Russian learner: the intonation patterns, being correct in their form, fail to convey the function, being wrongly applied to the situation.

For example, in the pronunciation of the English low-falling nuclear tone Russian students showed a tendency towards the contour of the Russian low-falling tone, being less gradual and less ‘final’ than the corresponding English tone. Russian students tend to pronounce the English low-rising nuclear tone in an abrupt manner, rather than a gradual one, typical of the English low rising melodical contour (Take-off). Complex and compound English nuclear tones (Fall-Rise, Fall+Rise, Rise-Fall) are especially difficult for Russian students. The whole diapason of these tones is much narrower and in the point at which there is a change of the tone it is levelled out.

Foreign accent can be viewed as a marked variation of speech constituting a unified entity and betraying the non-native origin of the speaker (in the case of a bi-lingual contact) which can be detected both in the segmental and the suprasegmental characteristics of oral speech. The ‘inferiority complex’ created by the notion of ‘foreign accent’ in the mind of the non-native speakers of English is a great communication hazard. It is limiting their ability to get their message across when conversing with the native speakers: they are afraid that the latter would misjudge them. There is a certain prosodic model of Russian accented speech containing
both universal and specific features of the Russian learner’s error performance. Russian accent in English intonation is marked by:

- wrong form and distribution of nuclear tones;
- distortion of the melodic contour in other parts of the melodic contour (prehead, head, tail);
- narrow voice range;
- low voice register in the onset of utterances;
- wrong distribution of prominence peaks within an utterance with a tendency towards stressing the final word;
- uneven rhythmical arrangement of utterances;
- monotony of tempo;
- wrong distribution of pauses and absence of variation in their duration;
- **Russian voice timbre**

The overall pattern of English intonation in the pronunciation of Russian learners: it is devoid of natural variation of intonation components and their constituents. The quality of the intonation components along with the irregular frequency of their appearance in the speech flow (e.g. type of nuclear tone – tone direction, range or register) creates the effect of the non-native accent. The wrong use of the intonation pattern (contradicting the pragmatics of communication) adds up to the overall impression of non-nativeness in oral speech.

It is acknowledged that communication problems arise when listeners pay more attention to the speaker’s accent (his/ her peculiar way of saying sounds and words accompanied by uncommon voice inflections) than to what he or she is saying. Accents can hamper performance, and may be a source of concern and embarrassment. On the job, a strong accent can adversely influence advancement and promotion: “Whether we like it or not, the ability to handle one of the most widely acceptable accents has become an important indicator of an individual’s ability to control the world around him. As such, it is a signal which must prove of great interest to an employer, and we know that accents play a significant role in influencing the often crucial first impressions formed at job interviews” (Honey 1989: 177). People say that only thick-skinned or very famous people can allow themselves the luxury of speaking with a heavy accent. On the other hand, a mellifluous accent can perform miracles — even in the sphere of politics (Op. cit.: 132).

The public, as the accent research shows, sometimes associates foreign accents with lower social status. And speakers with foreign accents may often be viewed as less intelligent, less competent, and even less attractive than native English speakers. These observations
underline the importance of a teacher’s phonetic instruction in ESL/ EFL teaching.

Intonation research is crucial for improving the pronunciation teaching and learning techniques of English as a foreign tongue, in the situation of classroom bilingualism. Adrian Underhill (1998: 194) rightly observes that intonation “…seems not to be accessible to direct cognitive intervention as the pronunciation of individual sounds or the manipulation of grammatical constructions or the learning of new vocabulary. Perhaps intonation is controlled by a different part of the brain, less accessible to conscious intervention”. The intonation of one’s native tongue is lying dormant in the brain, being very much subconscious, since one acquires it at the earliest period of life, even before one learns to say words. That is why automatic use of intonation patterns of a foreign tongue comes difficult to language learners who begin their language acquisition with learning words rather than with listening comprehension, the practice that every baby is exposed to in his/her native tongue the very minute he/she comes into this world. J. D. O’Connor very truly wrote (1967: 17): “English intonation is English: it is not the same as the intonation of any other language. Some people imagine that intonation is the same for all languages, but this is not true. You must learn the shapes of the English tunes, and these may be quite different from the normal tunes of your own language, and you must learn the meanings of the English tunes too, because they are important”.

The teaching of intonation remains to be less effective and more hazardous than the teaching of other phonetic aspects of a foreign tongue. There is still no adequate and learnable description of English intonation for Russian learners. There are substantial reasons for it. Firstly, it is due to the nature of intonation itself, being a very complex prosodic phenomenon. Secondly, existing descriptions of intonation seem to be incomplete. Thirdly, intonation is less tangible than other phonetic characteristics of oral speech. However, there are now new ways of improving suprasegmental qualities of oral speech by applying high technology to the teaching and learning of languages.

4. Conclusion

Russian linguists have always treated intonation as a complex unity of several suprasegmental features of oral speech: melody, sentence stress, tempo, and timbre. Each of these components is complex in its structure and character of realization: melody comprises the direction of voice pitch (falling/rising patterns), voice register and voice diapason, configuration of the tone, etc.; sentence stress may be normal, emphatic,
logical, centralized, decentralized, etc.; tempo includes tempo itself (slow, fast, normal), rhythm (with different rhythmical patterns) and pausation (pauses, different in their quantitative and qualitative characteristics); timbre is a complex of suprasegmental elements responsible for the overall voice quality effect (voice colouring).

The enormous bulk of intonation data collected from the study of the world languages ask for further scrutiny of this speech property, all the more so that intonation has become a focal theme of research not only in linguistics but in other intellectual spheres (literature, music, ballet, sculpture, theatre, etc.). Intonation is a language universal, and as such it is supposed to be treated like one. In one of the latest works on intonology in this country an idea was put forward about creating the concept of ‘universal intonology’ (Intonology 2006: 10). In his well known work “Stylistique francaise” Charles Bally (1912) wrote that intonation was not only a ‘speech commentator’ but also a ‘thought commentator’. At the dawn of the XXI century this postulation is being developed by linguists in the new approach towards intonation as a an integral part of the human thinking process. Intonation is said to be one of the most dynamic and functional concepts of the century. This concept of intonation no longer belongs to the purely linguistic sphere, it is now accepted as a basic notion by many other humanitarian and natural sciences.

One of the basic aims of phonological theory is to attempt to explain how and why languages differ from one another phonetically, by setting up a limited number of phonological parameters which can combine in various ways to generate the appropriate range of phonetic variability (Hirst 1976). Most recent studies on intonation show that there is still a lack of data about the permitted range of inter-language variation, and where descriptions of the intonation of different languages are available it is difficult to relate them to one another because of the differing modes of description adopted (Fox 1995). And it is not sufficient merely to register the differences in order to establish typological parameters in intonation (Op. cit.). A number of questions remain unanswered concerning both the way in which phonological representations of intonation structures are derived and the way in which those representations are interpreted phonetically. Since most intonation studies are heavily depending on empirical data, and not on phonological insight, intonation is still treated as a non-linguistic property — just as a feature of human voice which conveys emotions and attitudes (Martin 2000). However, different intonation contours, along with a variety of other suprasegmentals, are loaded with a range of linguistic meanings maintaining a rigorous system
of contrast with each other same as other phonological units (vowels, consonants) do.

The voice itself is a form of nonverbal communication providing a vital link with other people. The impact of intonation upon people is strong. In fact, research shows that listeners pay more attention to the vocal messages than to the words. “It is not what you said, it is how you said it” - is the phrase one often hears in everyday situations. The voice communicates through its tone, speed, pitch, volume, number, and length of pauses, and disfluencies (temporizers or fillers-in, of the type ‘uh’, ‘um’ ‘er’, etc.). All these factors can do a great deal to reinforce or contradict the message our words convey. Intonation is an essential part of our verbal and non-verbal communication. Other factors in intonation strongly influence our attitudes to other people. Among them is the factor of accented speech.

Foreign accent as a marked variation of oral speech constituting a unified entity and betraying the non-native origin of the speaker can be detected both in the segmental and the suprasegmental characteristics of oral speech, particularly in intonation (Vishnevskaya 1993). The study of interference effects resulting from the overlap of the two prosodic systems is indicative of the fact that accented speech is marked by an explicit prosodic model characterized by both universal and specific features of the foreign learner’s error performance. The «gammas» and «arpeggios» in the teaching of English suprasegmentals are indispensable to building up the natural prosodic continuity of speech.

For many years in Russia it was RP, as a British pronunciation standard pattern of educated speech, which was accepted as the teaching norm of English. However, the situation is changing now. Many students tend to pick up other accents (American accents, for instance, are often heard), and they have every right to do so in this changing globalised world of to-day.

While it is generally acknowledged that intonation contributes a lot to the impression of foreign accent, the number of studies and analyses explicitly dealing with the description of intonational foreign accent is rather small, particularly in the context of models of foreign language acquisition. Alan Cruttenden is right in prophesying (1986: 183) that the study of intonation is at a point from which there is likely to be significant and consistent progress in the next decade or two. It is true, that now it is “an exciting time for intonationists” (Op. cit.: 184). The beginning of the 21st century is characterized by immense interest in prosody and intonation research of different types of oral discourse, the semantics and pragmatics of intonation. The goal of phonetic research nowadays is to give a
description of intonation in a cross-cultural perspective, as a linguistic category possessing its own phonetic features and functions in oral native and non-native ‘lingua franca’. And last but not least: there remains the supreme task – to go from theory and research to practice.

References


CONTRIBUTORS

Tsvetanka Chernogorova works as a Senior Assistant Professor of English Language and English Phonetics in the Department of English and American Studies, University of Sofia, Bulgaria. She has taken part in teacher-training seminars for English teachers in Bulgaria, and participated in a number of conferences. The focus of her work is English phonetics, as well as English pronunciation models, English dialects, varieties of modern English.

Biljana Ćubrović 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). She was a Visiting Scholar at the postdoctoral level in the Department of Linguistics at Cornell University in 2008/2009. She is the author of A Workbook of English Phonology (2003, 2005), The Phonological Structure of Recent French Loanwords in Contemporary English (2005), and Profiling English Phonetics (2009), as well as a co-author of the series English Entrance Exam Practice – tests with key and explanatory notes. Dr. Ćubrović 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). She has given invited talks in Japan, Slovenia, UK and Serbia, taken part in teacher-training seminars for English teachers in Serbia, and participated in a number of international conferences. She was awarded a Certificate of Proficiency in the Phonetics of English by IPA in 2006. She is currently working for the TEMPUS project on the reform of foreign language studies in Serbia and is involved in a translation studies project approved by the Serbian Ministry of Education and Science for 2011-2014. The focus of Dr. Cubrović’s work is English phonetics and phonology, linguistic and cultural interplays between English and Serbian, language contacts, and EFL testing.

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 received her PhD in 1999 for the thesis
Contributors 242

Rhythmic Organization in English and in Bulgarian: a contrastive phonetic investigation. Some of the courses that she is currently teaching include English Phonology, Varieties of Spoken English, The Languages and Cultures of Scotland, Phonetics beyond the Basics, as well as a number of MA-level and teacher retraining courses in English phonetics. She spent the academic year 2005/2006 working as a Research Associate in the Department of Linguistics and English Language, University of Edinburgh. Her primary research interests include experimental phonetics and socio-phonetics, English and Bulgarian phonology, the prosody of English, the teaching of English pronunciation and speech corpora. She has presented at international conferences, has participated in several bilateral and multilateral research projects, and has published extensively both at home and abroad. She is the author of English Pronunciation for Bulgarians (2003).

Milena Insam is a lecturer of English linguistics at the Department of English Studies at Karl-Franzens University Graz (Austria), where she received her MA in 2009 (graduated with honours). She spent the academic year 2005/2006 at Rider University in Lawrenceville, NJ, and the summer term 2008 at the University of Florence (Italy) carrying out research for her Master's thesis in English literature. Ms. Insam is currently pursuing her Doctoral studies at Karl-Franzens University Graz writing her dissertation in the field of English linguistics with a focus on native and non-native pronunciation in English. She was awarded a doctoral scholarship by the Faculty of Arts and Humanities and received grants for research projects at the Public Library in New York and the University of Connecticut, Storrs (in 2009 and 2010, respectively), where she conducted extensive research and collected data for her dissertation project. Ms. Insam is a research assistant in the project 'Fauvoit', an interdepartmental venture in the field of FLT at Karl-Franzens University Graz. Within the course of the PhD program, she has participated in several international conferences and published articles on preliminary findings of her dissertation project. The focus of her work is English phonetics and phonology, language acquisition, and language teaching research.

Bojana Jakovljević is a PhD student at Novi Sad University (Serbia) where she received her MA degree in 2009. Her MA thesis, entitled Articulatory and Acoustic Description of Obstruents in English and Serbian, examines similarities and differences between English and Serbian obstruents, focusing on their acoustic features. Her subsequent
Klementina Jurančič Petek is Assistant Professor of English language and English phonetics in the Department of English and American Studies, University of Maribor (Slovenia). She is current Head of the same department. Dr. Jurančič Petek received her MA (1998) and PhD (2004) from the Faculty of Arts, University of Ljubljana (Slovenia). Her PhD thesis dealing with the phonetic and phonological characteristics of Slovene English was published in 2007 in a monograph entitled The Pronunciation of English in Slovenia: English Spoken by Slovene Learners, its Development and Factors Influencing it (ISBN: 978-961-6320-47-4, Maribor: Slavistično društvo). She is also one of the co-authors of the textbook package Touchstone 8 (ISBN 961-6239-38-4, Ljubljana: Tangram) for English in the 8th grade of Slovene elementary schools. Dr. Jurančič Petek has published professional and scientific articles relating to English phonetics and phonology, English linguistics, translation and history of art, both in Slovenia and abroad. She is currently teaching English phonetics and phonology, the English verb, British dialects, Modalities in English and Psycholinguistics.

Ken-Ichi Kadooka is Professor of English in the School of Business Administration, Ryukoku University (Kyoto, Japan). He received his MA degree at Kobe University in 1990. He is the author of Nihongo Onomatopeia Goi ni okeru Keitaiteki, On-inteki Taikeisei ni tsuite (On the Morphological and Phonological Systematicity of Japanese Onomatopeia Lexicon, Kuroshio Publisher, 2007) and Setsu-onchou no Kinou ni tsuite no Sentaku-Taikei Bumpou ni yoru Bunseki (A Systemic Analysis of the Clause Intonation, Seibido 2009). His focus of studies is the function and meaning of the intonation systems across languages, and Japanese onomatopoeia lexicon.

Anastazija Kirkova-Naskova is Teaching Assistant in English Phonetics and Morphology in the Department of English Language and Literature, Faculty of Philology, University of Skopje (Republic of Macedonia), where she received her MA degree in 2009 (MA thesis: Markers of foreign accent in Macedonian-accented English). She is currently working on her
PhD thesis: *The influence of perceptual training on the English pronunciation of Macedonian EFL learners*. In 2007 she attended the UCL Summer Course in English Phonetics (EFL Strand). Ms Kirkova-Naskova has published a number of research papers and reviews in scientific journals and book collections. She has presented her research at international conferences in France, Poland, Serbia, and Turkey, and also participated at conferences in Germany and the UK. She is also actively involved in two on-going research projects: *English Pronunciation Teaching in Europe Survey – EPTiES* (an on-line survey jointly prepared by teachers from 10 European countries), and *Corpus of Macedonian English Interlanguage* (a national project funded by the Ministry of Education and Science of the Republic of Macedonia). Her research interests focus on L2 phonological acquisition, L2 prosodic acquisition, linguistic perception, L1 interference and accent, and pronunciation instruction. She is also a trained UCLES oral examiner.

**Isabel Landsiedler** studied English and Spanish at the University of Graz, Austria, where she also wrote her PhD thesis on the semiotic aspects of creativity in advertising, after having studied abroad in Spain and England. She taught at a variety of institutions (University of Graz, Graz Technical University, University of Applied Sciences in Pinkafeld, Medical University in Graz), and from 2001 to 2008 she was the director of *treffpunkt sprachen*, the language centre of Graz University. She is currently employed as a senior lecturer at Graz Technical University and teaches linguistics at the Department of English Studies of Graz University. She participated in and coordinated several international, national and local projects focusing on different aspects of language learning (LES, UNIDISS, LAMM, TNP 2, ENLU, VUS, Language Server, LanQua, FauvoT etc.). She gave invited talks in Ireland, Slovakia, Germany, published several articles on language learning and has been member of the scientific board of the European Centre for Modern Languages of the Council of Europe (ECML) since 2007, where international teams work on innovative methodological approaches to language learning.

**Takehiko Makino** is Associate Professor of English as a Foreign Language at Chuo University (Tokyo, Japan) and has taught English phonetics at a number of other universities including his *alma mater* Tokyo University of Foreign Studies, from which he received an MA in English linguistics in 1991. He also studied linguistics at the University of Kansas, USA in 1987-88. In August 2004, he was awarded a Certificate of

**Maja Marković** is an Assistant Professor at the English Department of Novi Sad University, where she teaches courses in Phonetics and Phonology at BA and MA level. She specialized in English Phonetics and Phonology and Second language phonology. She received her MA and PhD degrees at Novi Sad University (in 1999 and 2007 respectively). Her MA thesis is an account of variable word stress in contemporary English. Her PhD thesis, titled *The Acoustic Analysis of English and Serbian Vowel Systems*, examines and compares the acoustic properties of vowels in English and Serbian, on the basis of a comprehensive empirical study. Her subsequent work has mainly been in the field of experimental phonetics, where she investigates the segmental and suprasegmental features of Serbian, the acquisition of the phonological system of English as a second language and the phonological systems of Serbian and English in contrast. She has participated in several national and international research projects.

**Brian Mott** is a lecturer in Linguistics in the English Department of the University of Barcelona, where he teaches Phonetics, 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). He also teaches on the Summer Course in English Phonetics (SCEP) at University College London with Professor John Wells and his colleagues. 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.

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. She received her PhD in 2003, for the thesis *Phonetic and phonological interference in the perception and production of English vowels by Serbian EFL students*. 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* (178002, Serbian Ministry of Education and Science), and the TEMPUS project aimed at the reform of foreign language studies in Serbia. Her research interests include sociophonetics and sociolinguistics, language attitudes, language varieties, and intercultural communication, particularly as related to applied linguistics and teaching English as a foreign/ additional language.

Ingrid Pfandl-Buchegger, Mag. Dr. phil., studied English, French and General and Applied Linguistics at the University of Graz, Austria, where she obtained a teacher's degree for English and French and later did her PhD thesis on the contemporary British novel, after having studied abroad at the University of Athens, Ohio, and the Ecole des Hautes Etudes en Sciences Sociales in Paris, France. She is currently employed as a senior lecturer at the Department of English Studies of Graz University, where she teaches English literature, pronunciation, and didactics. She is the director of the project FauvoT, a joint inter-departmental project in the field of foreign language teaching at the University of Graz that focuses on concentrated listening, in cooperation with the Technical University Graz and a neuropathic hospital, the Landesnervenklinik Sigmund Freud, where she also works as an academic advisor. Her current research explores aspects of foreign language learning, in particular the teaching of
pronunciation (she has published numerous articles in the field together with Eva Eberl, Isabel Landsiedler and Milena Insam) and intermedia studies (literature and dance, narrative across the media, metareferentiality in dance) in collaboration with a colleague from the University of Music and the Dramatic Arts in Graz, Austria, Gudrun Rottensteiner.

Vladimir Phillipov is Senior Assistant Professor in English Phonetics and Phonology in the Department of English and American Studies at the St. Kliment Ohridski University of Sofia. He graduated from the University of Malta (1982-1987). He spent three academic years as a Visiting Lector in Bulgarian Language and Literature at the universities of Leeds and Sheffield, UK (1992-1995), and another three years as a HESP Soros Foundation research worker at the Institute for Theoretical Linguistics, Budapest, Hungary (1996-1999). He delivered a course of lectures on English Intonation at the University of Chanakkale, Turkey (2001) and was a Visiting lecturer in linguistics for the winter semester’ 2006 at SUNY, Albany, USA. He participated in a number of international phora in phonetics: Barcelona, 1995; San Francisco, 1999; Budapest, 1999. The focus of Mr Phillipov’s work is intonation from a typological point of view and also related issues from the viewpoint of general linguistics.


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 spent a semester as 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, and English-Slovene Translation. He has also taught general linguistics and general phonology, and English Semantics and Pragmatics. 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.

**Dimitar Trajanov** is Associate Professor and Acting Dean of the Faculty of Computer Science and Engineering, Ss Cyril and Methodius University, Skopje, Republic of Macedonia. He received his MSc and PhD degrees in Computer Science and Engineering from the University of Skopje in 1998 and 2006 respectively. From 1999 to 2010 he was employed at the Faculty of Electrical Engineering and Information Technologies, first as research and teaching assistant, from 2006 as Assistant Professor and form 2010 as Associate Professor. From 2009 he is a Visiting Professor at the University of Stip, Republic of Macedonia. Dr Trajanov has published more than 75 journal and conference papers. He has been involved in 20 international and national scientific and applicative projects (including TEMPUS, FP6, COST, ERASMUS) as a project leader or a participant. He is currently President of the National Committee for ICT in Education of the Republic of Macedonia.

Galina M. Vishnevskaya is Professor of English Philology at Ivanovo State University (Russia). She received her Doctorate degree in the Theory of Language (following her PhD in Germanic languages) in 1993 (Leningrad State University, Russia). Dr. Vishnevskaya is Head of Ivanovo Phonetic Association since 1975. She has over 150 publications in the field of linguistics (in Russian and English) and is Editor-in-Chief of a series of papers in linguistics and phonetics. The focus of Dr. Vishnevskaya’s academic work: language contacts, language variation, bilingualism, foreign accent, English and Russian phonetics, pronunciation teaching.
# Index

## A
- acoustic analysis, 13, 14, 18, 34, 132, 154, 156
- alveolar, 37, 48, 54, 188, 199, 203
- American English, 57, 58, 59, 71, 109, 116, 117, 118, 119, 120, 122, 156, 170, 183, 192, 194, 211, 212, 213, 215, 217, 218, 219, 221, 224, 245
- amplitude, 18, 24
- aperiodic, 18, 19, 21, 28
- articulation, 3, 11, 14, 17, 21, 22, 24, 27, 28, 31, 33, 36, 37, 38, 39, 42, 44, 48, 117, 130, 133, 190, 198, 199, 200
- aspirated, 32, 33, 36, 44, 47, 48, 53, 54, 190, 202
- aspiration, 32, 37, 39, 40, 41, 43, 44, 45, 47, 50, 53, 54, 173, 188, 197, 198, 200
- assimilation, 14, 16, 21, 25, 27, 28, 30, 70, 140, 197, 198, 201
- attributive adjectives, 74, 76, 77
- audio-phonology theory, 126, 131

## B
- broad focus, 76
- Bulgarian, 99, 102, 107, 207, 208, 216, 217, 219, 220, 221, 222, 242, 247

## C
- casual style, 58
- Cockney, 3, 4, 5, 6, 7, 8, 10, 11, 215, 217, 227
- compression, 7
- connected speech, 57
- consonant clusters, 16, 17
- Contrastive Analysis Hypothesis, 32
- conversational speech, 58, 71
- *covert prestige*, 4
- Cruttenden, 10, 11, 16, 29, 58, 71, 104, 105, 108, 119, 122, 192, 198, 204, 205, 229, 238, 239
D

deletion, 7, 70
descriptive approach, 114
descriptivism, 113, 114, 115, 121
Deterding, 3, 11
devoicing, 14, 17, 20, 21, 24, 30, 36, 40, 44, 188, 189, 197, 198, 201
diachronic, 104
dialect, 75, 83, 165, 166, 170, 191, 246
diphthongs, 6
discourse, 74, 82, 85, 89, 99, 106, 108, 151, 152, 161, 183, 238
double burst, 48
duration, 18, 20, 24, 31, 33, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 49, 89, 150, 151, 152, 153, 183, 186, 197, 198, 235

E

EFL, 47, 53, 120, 145, 146, 150, 151, 152, 160, 208, 210, 222, 223, 236, 241, 244, 246
EFL learners, 47, 244
Electronic Ear, 129, 132, 141
ELF, 208, 210, 211, 222
ESL, 165, 170, 172, 174, 208, 236
Estuary English, 4, 227

F

FauvoT, 125, 131, 132, 134, 135, 136, 142, 244, 246
fieldwork, 45, 163
Flege, 163, 171, 172, 173, 176, 177, 180, 193

foreign accent, 180, 181, 185, 192, 194, 225, 228, 231, 232, 233, 234, 238, 243, 250
formal style, 57
formant frequencies, 9, 10, 18, 127, 135, 138
formant structure, 17, 19, 20, 21, 24, 25, 133, 138, 140, 150
formants, 11, 20, 23, 24, 27, 134, 135, 138, 140
fortition, 15, 28
free speech, 154, 163, 164, 165, 166, 167, 168, 171, 172, 173, 174, 176, 183, 191
French, 241
frequency, 8, 10, 19, 20, 21, 23, 75, 102, 117, 126, 127, 128, 129, 132, 134, 135, 138, 140, 151, 185, 186, 187, 188, 191, 235
friction, 16, 17, 19, 20, 22, 24, 25, 27, 28, 29

G

General American, 116, 166, 170, 183, 207, 209
glide insertion, 6, 7

H

HappY Tensing, 4
hiatus, 6, 7

I

informal style, 58
intensity, 19, 20, 23, 24, 74
interaction, 146, 155, 156, 157, 159, 181, 191
interference, 32, 132, 165, 166, 170, 171, 174, 183, 227, 238, 244, 246
interrogatives, 74, 75, 76, 77
interviews, 58, 62, 147, 156, 157, 235
intonation patterns, 88, 90, 92, 141, 211, 230, 234, 236
IPA, 13, 158, 241, 245
IPA symbol, 14

J

Japanese, 71, 73, 75, 82, 83, 86, 89, 90, 92, 94, 95, 96, 97, 208, 243, 245, 248, 249
joke, 85, 86, 87, 90
jokes, 85, 86, 88, 89, 90, 91, 92, 96

K

Kobanashi stories, 91, 92
Kuhl, 125, 129, 130, 142

L

labiodental approximant, 13
labiodental fricative, 13, 14, 30
Ladefoged, 15, 30, 32, 37, 45, 158, 245
Laver, 146, 158, 159, 160, 240
lenition, 15, 24, 28

LFC, 210
Lingua Franca, 207, 208, 210, 211, 214, 223
listening, 76, 116, 125, 126, 128, 129, 130, 131, 132, 134, 135, 138, 140, 141, 142, 143, 175, 179, 231, 236, 246
London Speech, 3, 4
low-prestige variety, 4
L-Vocalization, 4

M

Macedonian, 179, 181, 182, 183, 184, 187, 188, 189, 190, 191, 193, 194, 243
minimal pairs, 48, 171
mixed methodology, 148
modernized prescriptivism, 113, 114, 116, 117, 120, 121
monophthongs, 5, 6, 11

N

narrow approximant, 15
native speaker, 77, 157, 181, 182, 192, 208, 209, 214, 226, 228, 231, 232, 233, 234
nuclear stress, 73, 77, 78, 79, 81, 82, 249
O'Connor, 6, 11, 229, 236, 239, 240
obstruent, 14, 16, 17, 21, 188, 189, 202
Ohala, 28, 30, 71
Orthographic Depth Hypothesis, 174

paratone, 85, 87, 88, 90, 91, 92, 94, 97
pause, 18, 85, 86, 87, 88, 89, 90, 91, 92, 95, 96, 105, 174, 190
periodic, 18, 22, 27, 28
phonation, 18, 20, 23, 26, 27, 32, 35, 36, 39, 44
pitch, 19, 21, 49, 74, 75, 83, 86, 87, 100, 107, 127, 152, 153, 160, 229, 230, 236, 238
pitch accents, 49
pitch contour, 19, 21
plosive, 14, 50, 51, 52, 198, 199, 203
Polarization principle, 33
Praat, 18, 34, 49
pragmatic, 82, 86, 87, 93, 99, 151, 152, 156, 157, 164
prescriptive approach, 115
prescriptivism, 113, 114, 115, 116, 121, 123
prominence, 73, 74, 75, 76, 77, 229, 230, 233, 235
prominent, 32, 58, 74, 75, 78, 226
pronunciation model, 207, 209
pronunciation teaching, 125, 239, 244
Punch Line, 85, 87, 90, 91, 92, 94, 97
qualitative approach, 146
qualitative methods, 145, 146

Rakugo, 91, 92, 96
Received Pronunciation, 207, 209, 221, 227
rhyming slang, 4
RIP, 210
RP, 3, 4, 5, 6, 10, 11, 59, 116, 117, 118, 120, 170, 171, 204, 207, 209, 210, 215, 220, 221, 227, 238, 248
S
Saussure, 99, 101, 109
schwa, 6, 7, 8, 10
sign, 99, 101, 102, 104, 105, 106, 107, 108
Slovene, 14, 15, 30, 160, 163, 165, 170, 171, 173, 174, 175, 197, 199, 203, 204, 205, 243, 248
speech perception, 126, 129, 130, 135, 140, 142, 145, 146, 155, 192, 230, 243
speech production, 55, 126, 130, 142, 145, 146, 155
speech rate, 48, 100
stops, 21, 31, 32, 33, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 47, 48, 50, 53, 54, 55
suprasegmental, 100, 102, 115, 146, 154, 225, 228, 230, 233, 234, 236, 238, 245
syllable, 14, 16, 23, 27, 31, 35, 37, 77, 86, 87, 150, 173, 201, 233
synchronic, 105

T
talk-in-interaction, 155
tempo, 48, 49, 89, 90, 96, 100, 102, 153, 229, 235, 236
test types, 163, 164, 165, 168, 169, 170, 171, 172, 173, 176
T-glottalling, 4
Tomatis, 126, 127, 128, 129, 130, 131, 132, 143
tone group, 103, 104
tonic placement, 73, 74, 79, 80, 81, 82
transfer, 13, 27, 31, 83, 248

U
unaspirated, 32, 48, 202
usage-enriched descriptivism, 113, 114

V
varieties, 4, 5, 86, 118, 120, 166, 170, 180, 182, 208, 209, 227, 241, 245, 246
Voice Onset Time, 31, 32, 47, 55
voiced, 6, 13, 14, 15, 16, 18, 19, 20, 21, 23, 24, 25, 29, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 43, 44, 48, 168, 169, 200, 201, 202
voiceless, 14, 15, 16, 18, 21, 25, 26, 27, 31, 32, 33, 36, 37, 39, 40, 41, 42, 43, 44, 45, 47, 48, 49, 54, 168, 169, 173, 188, 190, 199, 200, 201, 202
voicing assimilation, 14
VOT, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 47, 48, 49, 50, 51, 52, 53, 54, 55
vowel, 3, 4, 5, 6, 7, 8, 9, 10, 11, 15, 19, 20, 23, 24, 32, 33, 34, 36, 39, 49, 63, 68, 104, 133, 134, 135, 136, 137, 138, 139, 140, 150, 151, 157, 160, 173, 188, 189, 197, 198, 201, 202, 203, 233
W

weak torms, 57, 66, 71
web technologies, 179
Wells, 3, 5, 6, 7, 8, 10, 11, 59, 71,
   75, 82, 83, 104, 109, 116, 119,
   120, 123, 209, 210, 223, 225,
   227, 228, 229, 240, 245
WSSE, 210

Y

Yod-Coalescence, 4