Proceedings of the
Phonetics Teaching and Learning Conference
UCL, London, 9–11 August 2017
CONTENTS
(Papers are in alphabetical order of first author’s surname)

About PTLC ........................................................................................................................... 7
Scientific Committee ............................................................................................................. 9

The expression of emotion in fairy tales: a multimodal approach to improve EFL students’ oral renderings
Lilián Ariztimuño .................................................................................................................... 11

Enhancing L2 learners’ perception and production of the Arabic emphatic sounds
Hajar Binasfour, Jane Setter, Erhan Aslan ................................................................................ 16

Questioning the teaching of “question intonation”: the case of classroom elicitations
Marina Noelia Cantarutti ........................................................................................................ 21

Technology – aided pronunciation teaching in an ESP/EAP course
Ana Cendoya ........................................................................................................................... 26

A corpus-based online Mandarin pronunciation learning system for Cantonese learners:
development, evaluation, and implementation
Hsuehchu Chen and Qianwen Han .......................................................................................... 30

Review of the Lingua Franca Core for English pronunciation teaching in Korea
Hyunsong Chung .................................................................................................................... 35

How imitation can help the acquisition of L2 pronunciation
Eva Estebas Vilaplana ............................................................................................................ 39

Intrusive tokens of aspiration in French learners’ L2 English
Christelle Exare ..................................................................................................................... 44

A genre approach to prosody: teaching intonation from a discourse perspective
Miriam Germani and Lucia Rivas .......................................................................................... 49

Reflecting upon students’ problems in phonemic dictations
Marina Grasso ....................................................................................................................... 54

LSC instruction on L2 pronunciation development
S.M. Mohibul Hasan .............................................................................................................. 59

Spectrogram reading as a tool to teach acoustic phonetics
Mirjam J.I. de Jonge ............................................................................................................... 63

Learner perspectives on pronunciation feedback
Pekka Lintunen, Aleksi Mäkilähde and Pauliina Peltonen ...................................................... 67
Identification of Mandarin high-level and high-falling tones by Vietnamese learners
Hui-Chuan Liu ............................................................................................................ 72

Sequencing and technology-aided activities in the acquisition of foreign sounds
Daniela Martino ........................................................................................................ 76

Using electropalatography in second language pronunciation instruction: a preliminary examination of voiceless German fricatives
Shawn L. Nissen, Kate E. Lester, Laura Catharine Smith, Lisa D. Isaacson and Teresa R. Bell ........... 80

L2 English speech rhythm of Japanese speakers: an alternative implementation of the Varco metrics
Yumi Ozaki, Kakeru Yazawa and Mariko Kondo ........................................................ 84

What does the rapid spread of /u/-fronting in American English have to do with the teaching of French phonetics?
Gladys Saunders ........................................................................................................ 89

Research-led teaching in phonetics: an exercise in research literacy
Jane Setter ................................................................................................................. 94

A brief intensive method to help Japanese learners perform English tonicity
Yusuke Shibata, Masaki Taniguchi, Young Shin Kim .................................................... 99

Perception of the General British vowels /ʌ/, /ɑː/, /ɒ/, and /ɔː/ by Slovenian speakers of English
Andrej Stopar ............................................................................................................. 104

Online video assessment of clinical phonetic transcription skills in Speech and Language Pathology
Claire Timmins ......................................................................................................... 110

Online video assessment of clinical phonetic transcription skills in Speech and Language Pathology
Kakeru Yazawa, Mariko Kondo and Paola Escudero ..................................................... 115
About PTLC

The Phonetics Teaching and Learning Conference was initially set up as one of the deliverables of a project funded by the Higher Education Funding Council for England and the Department of Education of Northern Ireland. The project, which ran from 1997 to 2000, had UCL as its lead institution and partner institutions at the universities of Cambridge, Central England, Newcastle, Ulster, Westminster and York.

The first meeting of PTLC was at UCL in 1999, and gave rise to a biennial series of conferences which built upon and expanded the original successful formula. Over the first 16 years researchers from a wide range of countries contributed no fewer than 147 papers, all of which were double-blind reviewed by members of a scientific committee before acceptance and publication. As our twentieth anniversary approaches, these present proceedings of PTLC2017 add a further 23 papers, pushing the tally to 170.

Following a pattern established at the 2015 conference, the programme for 2017 also included 2 invited talks on diverse topics to emphasise PTLC’s broad remit. Dr Dominic Watt, University of York, delivered a talk with the title ‘Private Ear training: Phonetics teaching for the next generation of forensic speech scientists’, while a whole morning session of the meeting was devoted to a double-length invited talk from Professor James E. Flege, University of Alabama, with the title ‘The cross-language acquisition of stops differing in VOT: Historical review and key findings’. This included a valuable review of all Professor Flege’s work in this area, and in an unusual step, pdf files of the slides from his presentations are being published online alongside these Proceedings. We are very grateful to Professor Flege for making this valuable resource available.
Scientific committee PTLC2017:

Mercedes Cabrera Abreu: Universidad de Las Palmas de Gran Canaria
Michael Ashby: UCL
Patricia Ashby: University of Westminster
Marina Cantarutti: University of York
Paul Carley: University of Leiden
Eva Estebas: Universidad Nacional de Educación a Distancia
Ulrike Gut: University of Münster
Valerie Hazan: UCL
Smiljana Komar: University of Ljubljana
Rachael-Anne Knight: City University
María Luisa García Lecumberri: University of the Basque Country
John Maidment: UCL (Chair)
Inger Mees: Copenhagen Business School
Francisco Vizcaino Ortega: Universidad de Las Palmas de Gran Canaria
Joanna Przedlacka: UCL and University of Oxford
Alex Rotatori, Università degli Studi di Tor Vergata, and Collegio IPASVI, Rome
Hiroko Saito: Tokyo University of Foreign Studies
Gladys Saunders: University of Virginia
Jane Setter: University of Reading
Masaki Taniguchi: Kochi University
J. C. Wells: UCL
The expression of emotion in fairy tales: A multimodal approach to improve EFL students’ oral renderings
Lilián Inés Ariztimuño

Universidad Nacional de La Pampa (UNLPam)
lilianarizti@humanas.unlpam.edu.ar

ABSTRACT
Effective oral renderings of stories involve the use of appropriate vocal cues to express emotions. This, however, represents a challenge for EFL teacher training students who do not only need to interpret emotions construed in the written text but also associate those emotions with certain prosodic and paralinguistic features. Therefore, awareness of the association between the written and oral expression of emotion constitutes a key issue in the EFL field. Nevertheless, its treatment remains sporadic and intuitive as no specific framework is available for a systematic teaching approach.

This paper presents the findings of an exploratory study intended to describe the association between the written and oral expression of emotion in 3 fairy tales: first, a preliminary taxonomy of phonetic profiles for emotions classified drawing on the system of APPRAISAL – AFFECT [19]; and second, the potential benefits of a multimodal teaching approach aimed at improving the effective expression of emotion of EFL students when reading stories aloud.

Keywords: expression of emotion, fairy tales, AFFECT, phonetic profiles, teacher training centres.

1. INTRODUCTION
Stories have been widely acknowledged as one of the primary means by which English-speaking cultures interpret life, evaluate behaviour and, most importantly within the EFL context, educate and entertain people of all ages [18]. The main point of narratives and thus fairy tales is to show “how the protagonists resolve a complication in their lives, once they have evaluated the complicating action with some type of attitude” [18]. Consequently, the identification, interpretation and description of the written and oral resources that realize the expression of attitudes and emotions as well as of the association of these resources are essential to understand stories and to give voice to them.

Although it could be said that a story may have as many interpretations as people reading it, it is equally true to state that there is what [23] call “a dominant reading”. For purposes of analysis of such an elusive subject as emotions, the present study assumes that the generic structure of narratives along with their evaluative meanings suggest “an ideal reading, a position from which characters and events become intelligible, values shareable and the narrative itself coherent” [15]. Within the Systemic Functional Linguistics (SFL) tradition several studies [17, 18, 23] have established the constitutive presence of evaluation in this genre and many researchers [4, 15, 16, 18, 26] have used APPRAISAL as an effective linguistic tool to describe and explain how language resources encode attitudes and emotions.

The oral realization of emotion has also been mentioned in relation to evaluation by a group of studies [9, 10, 11, 12, 14, 19, 26] but a systematic description of how, or if, it relates to the written linguistic resources has not yet been explored in depth. The largest group of studies devoted to emotional speech, however, have been carried out in the fields of computational linguistics and speech synthesis [8, 13, 21, 22, 24].

The importance of teaching emotional speech in contexts of second language acquisition has been established by Pavlenko in her book Emotions and multilingualism [20]. Two crucial conclusions for the topic of emotions from a cross-linguistic perspective are mentioned in her review: the fact “that linguistic background affects interpretation and expression of emotion through vocal cues in intercultural communication” and the “important implications for FL or L2 classrooms, where neither verbal nor vocal aspects of emotional expression have been getting much attention”. In the more local context of this study – EFL teacher training programs in Argentina – even though classroom experience shows that the expression of emotion is assigned an essential role in effective reading aloud and storytelling, no specific framework is available at present for a systematic EFL teaching approach. This becomes evident by the reduced number of studies on the teaching of the expression of emotion in its written and oral manifestation in Argentinian EFL contexts [1, 2, 6, 7].

2. THE STUDY
The study reported here offers the first findings intended to fill this gap by exploring the association between the written and oral expression of emotion.
in 3 fairy tales drawing upon the system of APPRAISAL – AFFECT [19], and an adaptation and simplification of the phonetic taxonomy for the transcription of emotional speech given in [22]. The educational goal of the study was to propose a multimodal approach designed to improve EFL students’ interpretation of the written evaluative meanings as well as their oral renderings of stories by means of relevant prosodic and paralinguistic features – emotional speech.

2.1 Theoretical foundations

The system of APPRAISAL [19] offers a theoretical framework within SFL which, focusing on the interpersonal function of language, systematizes the great variety of linguistic resources deployed by language users to express attitudinal meanings. The system of ATTITUDE focuses on “feelings, including emotional reactions, judgements of behaviour and evaluation of things” [19]. Out of these three regions developed within ATTITUDE, this study concentrated on AFFECT (the domain of emotions). To be specific, the corpus was analysed in terms of INSCRIBED AFFECT, i.e. all those linguistic expressions that denote emotions as classified in [4]’s most recent modifications to the subsystem of AFFECT. Within AFFECT, emotions can be grouped into five subsystems: un/happiness (“affairs of the heart” [19]); in/security (“ecosocial wellbeing” [19]); dis/satisfaction (pursuit of goals); surprise (“linguistic domain of unexpectedness” [16]); and dis/inclination (desire and non-desire). It is important to mention that the labels proposed by AFFECT encompass the great majority of words typically used to describe emotions from a psychological perspective. For example, if the well-known five universal emotions suggested by Ekman’s “Atlas of Emotions” [25] are considered — enjoyment, sadness, anger, disgust and fear, it could be said, for example, that enjoyment is labelled (according to AFFECT) as happiness: cheer, sadness as unhappiness: misery, anger as dissatisfaction: displeasure and fear as insecurity: disquiet. This fact shortens the methodological distance between this study and previous research on vocal cues and thus makes a comparison with previous work on emotional speech possible.

There is general agreement on the fact that changes in the voice can indicate affective meaning, but also on the many challenges faced when a systematic description of those cues is attempted [8, 13, 21, 22, 24]. One of these difficulties derives from the inconsistencies and number of options observed in the categories used to label emotions. The importance of reducing “emotion labels to a practically feasible number, without obscuring important differences between states, is a challenge for researchers of vocal affect” [13]. This work attempts to contribute to the existing research combining phonetics and the solid theoretical foundations of SFL for the analysis and classification of emotion terms.

For the purpose of the present study, the phonetic realization of emotion was described following [22]’s phonetic taxonomy. This comprehensive set of prosodic and paralinguistic features was originally designed with a computational objective in mind, thus the need to adapt and simplify it for the EFL context of this study. The description of all the prosodic and paralinguistic features proposed by [22] can be overwhelming for EFL students to deal with. Therefore, the phonetic taxonomy applied here was circumscribed to four prosodic features: pitch height, pitch range, loudness and tempo; and the paralinguistic features of voice quality, vocal effect and voice qualification. The rationale for the selection of this limited set of prosodic features can be justified with perceptual, acoustic and pedagogical arguments. Perceptually, changes in pitch (height and range), loudness and tempo are usually easily perceived by the human ear. These perceptual differences can be measured in terms of frequency, amplitude and seconds using acoustic software such as Praat [5] in order to corroborate listeners’ impressions. EFL teacher training students are usually trained in perceptual recognition of the prosodic and paralinguistic features analysed here and thus can understand and apply the phonetic taxonomy proposed.

2.2 Research question

This study had one main research question in mind:

1. How does INSCRIBED AFFECT in the stories analysed relate to the prosodic and paralinguistic features selected by the storyteller to express those emotions orally?

2.3 Methodology and results

The present study implemented an exploratory mixed methods design as it explored, described and related the written and oral expression of emotion in a corpus of 3 stories (King Thrushbeard, The elves and the shoemaker, Rumpelstiltskin) written by the Brothers Grimm and read aloud in British English by a professional female storyteller1. The analysis conducted was both qualitative and quantitative. Emotion terms (n=45) were first manually labelled for INSCRIBED AFFECT following the classification described in the theoretical foundations. Co-text, context and cultural background were taken into
consideration to decide which evaluative meanings spread through each clause containing the emotion terms. In order to validate the accuracy of the interpretation and labelling of the lexis in terms of AFFECT, a colleague trained on the use of the APPRAISAL framework was asked to check the researcher’s data analysis. All differences were discussed and settled so as to further validate this qualitative phase of the study.

Once the written expression of emotion was labelled, the clauses containing emotion terms were extracted from the complete audio file using Praat. Each clause was played as many times as necessary for the perceptual classification: first, to divide the clauses into tone units\(^2\) (n=88) and then, to value each tone unit in terms of the pitch height, pitch range, loudness, tempo and paralinguistic features used by the storyteller. Additional quantitative information (acoustic numerical measures) was obtained for the prosodic features so as to verify perceptual results.

2.3.1 Pitch height analysis

Relevant pitch height of the nucleus\(^3\) was considered for this analysis. It was classified perceptually as high, mid or low and measured with Praat in Hz using the command move cursor to maximum pitch.

2.3.2 Pitch range analysis

It was associated to the width of movement perceived on the nucleus of the tone unit. It was classified perceptually as wide, narrow or unmarked and measured in Hz by selecting the nucleus and using Praat’s commands move cursor to minimum pitch and move cursor to maximum pitch to get the highest and lowest values of pitch so as to then calculate the range between them.

2.3.3 Loudness analysis

It was perceptually classified as loud, moderate or soft and measured in dB using tone unit boundaries as limits for the audio selection. The Praat command get intensity was used to measure amplitude.

2.3.4 Tempo analysis

It was classified perceptually as fast, moderate or slow using the tone unit as the unit of analysis. It was measured with Praat counting the number of syllables per second (s). The categories proposed by [3] slow: 3.3 syllables per (s); medium: 4.3; and fast: 5.9 were taken as point of reference.

The accuracy and reliability of the data analysis was corroborated by an expert on phonetics and phonology. This external rater analysed the data following the same method conducted by the researcher. The results obtained by the researcher and the external rater were then compared reaching a 97% agreement. Differences were re-evaluated by both of them together so as to agree on the definitive results.

Perceptual qualitative results were then transformed into quantitative information as percentages were calculated for the frequency of occurrence of prosodic values for each AFFECT subtype. Table 1 displays the percentages obtained.

<table>
<thead>
<tr>
<th>Table 1 - Relation between AFFECT &amp; perceptual phonetic features</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFFECT subtype</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>dissatisfaction:</td>
</tr>
<tr>
<td>displeasure</td>
</tr>
<tr>
<td>n=21 tone units</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>insecurity:</td>
</tr>
<tr>
<td>disquiet</td>
</tr>
<tr>
<td>n=21 tone units</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>happiness:</td>
</tr>
<tr>
<td>cheer</td>
</tr>
<tr>
<td>n=20 tone units</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>satisfaction:</td>
</tr>
<tr>
<td>pleasure</td>
</tr>
<tr>
<td>n=13 tone units</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>unhappiness:</td>
</tr>
<tr>
<td>misery</td>
</tr>
<tr>
<td>n=13 tone units</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Note: H=high, M=mid, L=low, W=wide, Un=Unmarked, N=narrow, Ld=loud, Mod=moderate, S=soft, Fa=fast, Sl=slow. Main tendencies in **boldface** type.

Perceptual results and acoustic numerical measures obtained for the prosodic features were then compared so as to further validate the values suggested for each AFFECT subtype phonetic profile. Figure 1 shows the comparison between the perceptual and acoustic results for pitch height as an example.
3. DISCUSSION AND PEDAGOGICAL IMPLICATIONS

3.1 INSCRIBED AFFECT phonetic profiles

On the basis of the results obtained the association between the written and oral expression of emotion was described in the form of a provisional taxonomy of phonetic profiles for five INSCRIBED AFFECT subtypes: unhappiness, dis/satisfaction and insecurity. Table 2 displays these profiles.

<table>
<thead>
<tr>
<th>AFFECT subtype</th>
<th>pitch height</th>
<th>pitch range</th>
<th>loudness</th>
<th>tempo</th>
<th>paralinguistic features</th>
</tr>
</thead>
<tbody>
<tr>
<td>dissatisfaction: displeasure</td>
<td>high</td>
<td>wide</td>
<td>loud</td>
<td>fast to moderate</td>
<td>glottal attack, rough, creak, breathy, breath-in</td>
</tr>
<tr>
<td>satisfaction: pleasure</td>
<td>high to mid</td>
<td>wide</td>
<td>moderate</td>
<td>to fast</td>
<td>glottal attack</td>
</tr>
<tr>
<td>unhappiness: misery</td>
<td>low</td>
<td>narrow</td>
<td>moderate</td>
<td>to moderate</td>
<td>glottal attack, rough, breathy \ and turbulent</td>
</tr>
<tr>
<td>happiness: cheer</td>
<td>mid to high</td>
<td>narrow</td>
<td>moderate</td>
<td>to loud</td>
<td>creak, breathy, breathy-in, tremulous</td>
</tr>
<tr>
<td>insecurity: disquiet</td>
<td>low</td>
<td>narrow</td>
<td>soft</td>
<td>fast</td>
<td>glottal attack</td>
</tr>
</tbody>
</table>

The phonetic values assigned to each AFFECT subtype represent the interpretation of the frequencies and percentages obtained for each perceptual value (See Table 1) corroborated by the acoustic measures (See Figure 1) as well as the enumeration of paralinguistic features observed in the sample. For example, if we concentrate on unhappiness: misery, traditionally called sadness, we can see that pitch height was assigned the value ‘low’ in Table 2. This value represents the interpretation of the 100% assigned to ‘low’ for this AFFECT subtype as regards perceptual pitch height (See Table 1) which was corroborated by the acoustic measure mean of 227 Hz obtained for unhappiness: misery (See Figure 1). This number of Hz corresponds to those expected for low pitch height for a female speaker. Previous research on conversational situations has characterized sadness [13, 20] as displaying low pitch level, narrow pitch range, soft volume and slow tempo. Most results from the study reported in this paper are consistent with previous work. Differences in tempo could be explained in terms of the genres analysed: fairy tales and conversations.

The profiles displayed in Table 2 are not to be considered definitive, with exclusive and excluding features but rather open to future reconsideration and validation. Provisional as it is, however, this taxonomy offers the first attempt to describe the oral realization of INSCRIBED AFFECT. It presents the first findings resulting from the association of written linguistic resources and oral non-linguistic realizations of emotion in fairy tales.

3.2 Pedagogical implications: multimodal approach to deal with emotion in fairy tales

Based on the generalizations obtained, pedagogical implications for the teaching of the expression of emotion when reading aloud fairy tales at university level EFL teacher training programs will be suggested. The strongest implication that these findings put forward is the benefits a systematic multimodal approach is likely to have in EFL contexts. An approach that interprets meanings considering the context, co-text, lexical content of wordings and the phonetic realization of texts by competent readers will most certainly enhance the EFL teaching practice context.

This paper suggests a possible implementation of this approach which has rendered positive pilot classroom experiences, which, unfortunately, have not yet been documented. First, the EFL teacher familiarizes students with the basic system of AFFECT as a linguistic tool to uncover attitudinal meanings considering emotion terms, co-text and context. Then, the analysis of the written expression of emotion is combined with a phonetic analysis of a competent speaker’s oral rendering using the phonetic profiles for INSCRIBED AFFECT subtypes (See Table 2) for a better understanding of the interpersonal attitudinal meanings construed in the text. Finally, students are encouraged to voice the attitudes and emotions construed in the written text making effective use of prosodic and paralinguistic features.

EFL students’ oral production of emotion when reading aloud stories is likely to improve if students feel more confident when interpreting the attitudinal meanings of the written text. An explicit understanding of the different resources at play in the oral rendering of a text can make this happen.

4. REFERENCES


1 The stories were downloaded as text and mp3 files from the webpage http://storynory.com. Even though the webpage allows for free download, permission was granted by its editor and producer, Hugh Fraser, to use the material for this study.

2 This study followed the British tradition to define the term tone unit and thus, considered major pitch changes as signal of tone unit boundaries. Coincidence of tone unit and syntactic boundaries was favoured.

3 Even though relevant prosodic features related to pitch occur on the whole of the tone unit, only the nucleus was analysed as it is “the part given most emphasis by the speaker” [27] and thus most easily perceived by listeners.
ENHANCING L2 LEARNERS’ PERCEPTION AND PRODUCTION OF THE ARABIC EMPHATIC SOUNDS

Hajar Binasfour, Jane Setter, Erhan Aslan

University of Reading

h.s.m.binasfour@pgr.reading.ac.uk, j.e.setter@reading.ac.uk, erhan.aslan@reading.ac.uk

ABSTRACT

This study examined the Arabic L2 learners’ ability to perceive and produce the emphatic sounds /sˤ/, /ðˤ/, /dˤ/, and /tˤ/. Specifically, the study explored the effects of traditional-based and technology-based instruction in enhancing learners’ perception and production of these sounds. Data were collected from forced-choice identification tasks and audio recordings taken during pre- and post-test conditions. The results revealed that the emphatic sounds posed a considerable amount of perception and production difficulties to L2 learners of Arabic. Additionally, there were significant improvements among all participants after the traditional and technological training courses and that the difference in the outcome between the two teaching methods was not significant.

Keywords: Pronunciation instruction, Arabic sounds, Emphatics, Praat.

1. INTRODUCTION

The Arabic language is distinguished by the existence of four emphatic sounds /sˤ/, /ðˤ/, /dˤ/, and /tˤ/ with a primary articulation in the interior vocal tract and a secondary articulation in the pharynx. Emphatics are considered to be a unique characteristic of Arabic, while the absence of these sounds in most world languages results in pronunciation difficulties among L2 learners of Arabic [11, 15, 18]. The primary reason for these challenges in pronunciation is because of the acoustic and auditory similarities to their plain counterparts /s/, /ð/, /d/, and /t/, which exist in most languages [2, 24]. What distinguishes the emphatics from the non-emphatics is the effect of the former sounds on the following and preceding vowels causing an ‘emphasis or pharyngealization spread’ and altering these vowels to allophones [22, 26].

A number of studies discussed the significant similarities between emphatic sounds and their counterparts and how these sounds share similar acoustic features [4, 11]. These studies provided detail about the way emphatics and non-emphatics are articulated by Arabic native speakers. To date, very few studies discussed Arabic pronunciation and singled out the features of the emphatic sounds and the adjacent pharyngealized vowels as particular issues in teaching L2 Arabic pronunciation. The lack of knowledge and understanding about the differences between emphatics and non-emphatics among L2 learners of Arabic can cause perception and production difficulties [3, 6].

Many studies that investigated the role of phonetic instruction in L2 pronunciation teaching found a positive relationship between explicit instruction and the performance of L2 learners [20, 21]. For example, reading aloud, minimal pairs, repetition, and explicit phonetic instruction techniques revealed significant and positive results [13, 25]. Similarly, speech analysis technology alone with or without verbal phonetic instruction was found to lead to significantly improved pronunciation [14, 16, 19, 20]. Particularly, speech analysis technology is one of the modern tools that has been repeatedly tested and applied in teaching English segmentals and suprasegmentals [19, 20].

Speech analysis programs are used to create graphic representations of speech, which are based on the visual display of the articulation. The work on speech analysis technology in teaching pronunciation started in the late 1970s with a software called Visi-Pitch [10]. The creation of this software allowed researchers to investigate the potential benefits of teaching pronunciation through visual analysis of native speakers' speech [1, 7, 8, 27]. Praat, used in the present study, is an open-source speech analysis tool that is developed with manifold functions to help researchers analyse, measure and understand acoustic features of sounds [9] and shows visual movements of speech through waveforms and spectrograms.

The present study looked at the differences in efficacy between the traditional and the modern teaching methods using speech analysis software (Praat) in enhancing the perception and production of emphatics in L2 learner of Arabic. The purpose of the study is to see whether using visual representations of the acoustic features of sounds rather than the usual traditional teaching approach could help L2 learners in understanding the features of the emphatic sounds and hence perceive and produce them more accurately.
2. METHODOLOGY

An experiment was conducted to investigate the effect of two different teaching approaches in enhancing the perception and production abilities of emphatic sounds in L2 learners of Arabic.

2.1 Participants

For this study, 38 females L2 learners of Arabic from Princess Nourah University in Saudi Arabia participated in the training courses. The participants’ age was ranging from 20 to 26 years old (mean age 22.50), and they were from elementary, intermediate and advanced level of Arabic proficiency. 14 Urdu speakers, 13 Mandarin speakers, and 11 English speakers volunteered to participate. The time they spent studying Arabic ranged from three months to more than three years.

Participants were divided into two groups, 19 students in each group (see Table 1). They were divided equally based on their language backgrounds and proficiency levels in an attempt to control the effect of these variables on the results.

Table 1: Number of participants in each group (A= Traditional, B= Technology).

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Mandarin</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Urdu</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>English</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

2.2 Materials

Two sets of materials were designed for each group, one using speech software and the other using traditional methods. The materials for both groups contained similar words that were taken from an Arabic language learning series [5]. Special DMDX written scripts were also designed for the computerized perception and production tests.

2.2.1 Traditional group

The materials for this group included handouts which contained information about the manner and place of articulation of each emphatic sound; pictures of the vocal tract; example sentences; small passages; and minimal pairs.

2.2.2 Technology group

An introductory presentation for this group was designed using computer slide presentation software. For the training course, slides were designed for each of the four days. Each slide contained three waveform and spectrogram pictures of three syllables that have the same emphatic sound but in three different environments (e.g. /asˤaː/, /usˤuː/, /isˤiː/). Four sound files were prepared for this group which contained words and isolated syllables that have the emphatic and non-emphatic sounds pronounced by an Arabic native speaker.

2.2.3 Perception and production testing materials

The perception and production tests were administered to both groups by means of the DMDX Display Software which employed scripts designed specifically for this study. It presents stimuli materials and record participants’ responses via keyboard input [12]. Each test included thirty Arabic words placed in a carrier phrase presented in three phrases for each of the eight emphatics and non-emphatics, and six phrases which served as distractors. The words in the tests were minimal pairs to test participants’ abilities to discriminate between sound contrasts. The stimuli for all perception and production tests were similar but they were in random orders and they were not part of the set of words that was included in the training.

2.3 Procedure

This study took place in the Arabic language institute at Princess Nourah University. On the first day of the experiment, the demographic information questionnaires and ethical consent forms were distributed to the participants. After filling out the questionnaires, participants took the perception pre-test. They individually entered a quiet room and sat facing a laptop. They were asked to wear headphones and follow the instructions on the screen. On the following day, they took the production test in the same way.

After taking the pre-tests, all participants received instructions about the time and the place of the training. While the traditional group took it in a regular classroom and the technology group took it in a computer lab.

2.3.1 Traditional group

The training started with the introduction of the emphatic sound, its place and manner of articulation. A picture of the vocal tract was provided to explain the place of articulation and to show the position of the tongue. After that, participants practiced reading minimal pairs from the board and discriminate between the emphatics and non-emphatics in pronunciation such as: /maːsˤah/ ‘table’ and /maːsˤah/ ‘diamond’. /ʔaːbɪsˤ/ ‘stamp’ and /ʔaːbɪsˤ/ ‘follow’.
This training focused on the emphatic sound itself rather than the adjacent vowels. Participants were asked to read aloud sentences and a short passage individually in turn, and feedback was provided by the tutor when necessary. They spent one and a half hours on each of four days reading passages and sentences aloud, discriminating minimal pairs, and receiving verbal pronunciation instruction and feedback.

2.3.2 Technology group

The training for the technology group started with an introductory session about analysing sounds through Praat. Participants were given instructions with pictures about downloading and installing the software along with creating, opening and understanding spectrograms.

This group took the training sessions on the same days as the traditional group. The participants started by examining the features of the emphatic and non-emphatic sounds through Praat. The purpose was to teach the learners how to examine spectrograms and distinguish emphatics from non-emphatics. They then followed three steps as recommended by Offerman and Olson [19, 20]: initial self-recording, guided visual analysis and practice and re-recording, as outlined below.

In the initial self-recording stage, three syllables in isolation and three words were given to this group in each of the four days (e.g., /sˤa/, /sˤu/, /qasˤad/, /nusˤub/, /sˤiːn/). Participants were asked to record their voices through Praat then edit the recording to see the spectrogram and waveform.

In the guided visual analysis, the sound files of a native speaker pronouncing the same words and syllables were provided to participants (see Figure 1).

![Figure 1: Praat screenshot showing differences in the vowel formants between the emphatics /ðˤ/ and non-emphatic /ð/ in a native speaker of Arabic.](image)

The participants compared the shape of the emphatic consonants and the adjacent vowels in their spectrograms and those of the native speaker. The lowering of the second formant in the adjacent vowels was explained to participants at this stage. To enhance participants’ understanding, pictures of the vocal tract were provided to explain the articulation of these sounds and to justify the lowering of the second formants (F2).

In the practice and re-recording stage the participants re-recorded the required words again to compare them with the native speakers’ spectrograms. This allowed the participants to imitate the pronunciation of the native speaker many times, receive immediate feedback, and recognize the differences between the emphatic and non-emphatic sounds and adjacent vowels.

This group spent one and a half hours on each of the four days recording their voices, comparing them with native speakers’ voices, receiving immediate feedback many times, and imitating native speakers’ utterances.

2.4 Data collection

The perception data was taken from forced choice identification pre- and post-tests. Thirteen audio files of different phrases were played in random order to each participant. Two words appeared on the screen synchronizing with each audio file. Participants had unlimited time to think and decide which word they thought they heard.

The data taken from the production pre- and post-tests were audio recordings. 30 phrases appeared on the screen in random order and the participants were told to read them in a clear and loud voice. An Edirol R-09HR recorder was used to collect the data, the recordings sampled at a rate of 44100 Hz, 16 bit. The raters of these recordings were eleven Saudi Arabian language instructors who worked in different secondary schools in Saudi Arabia. Their task was to listen to the recordings and identify the incorrect sounds.

2.5 Data analysis

All data was analysed quantitatively using SPSS. A correct pronunciation for the production test and choice for the perception test were coded as (0, i.e., no error), and an incorrect pronunciation and choice were coded as (1).

Intraclass correlation coefficient (ICC) was conducted to calculate the inter-rater reliability of the production test raters. The results of the ICC showed a high degree of reliability between raters measurements. The average measure of ICC was rICC = .981 with a 95% confidence interval (α = 0.05) from .971 to .989, F (570.1) = 37.

A one-way between subjects (ANOVA) was conducted to compare between the two groups. Furthermore, a paired samples t-test was carried out to reveal whether any group improved significantly after receiving the explicit phonetic instruction.
3. RESULTS AND DISCUSSION

The results of the pre-tests revealed that L2 learners of Arabic faced great difficulties in the perception and production of the emphatic sounds, especially with the sounds /veis/ and /vifs/. Many participants from all the three proficiency levels produced errors in perceiving and producing the emphatics.

Based on the results of a paired samples t-test, significant improvements were found in the traditional group; \( t(18)=5.62, p < 0.001 \) and the technology group; \( t(18)=7.91, p < 0.001 \) after the training in the perception of the target sounds. However, the comparison between the two groups showed that there were no significant differences between the technology group (M=2.79, SD=2.2) and the traditional group (M=3.74, SD=2.6) at the \( p<.05 \) level in the perception of the emphatic sounds; \( [F(1, 36) = 1.513, p = 0.227] \).

Concerning production, significant improvements were found in the traditional group; \( t(18)=7.56, p < 0.001 \) and the technology group; \( t(18)=8.95, p < 0.001 \) after the training course in producing the target sounds. However, there were also no significant differences between the technology teaching group (M=2.42, SD=1.98) and the traditional teaching group (M=3.32, SD=2.3) in the production of the emphatic sounds; \( [F(1, 36) = 1.71, p = 0.200] \) (see Figure 2).

![Figure 2: Participants’ perception and production errors before and after receiving the two teaching methods.](image)

The statistical results showed that the technology group made fewer errors in perceiving and pronouncing the emphatics than the traditional group, but the difference between groups was not significant.

Both teaching approaches contributed significantly in developing L2 learners’ pronunciation of sounds. The results of this study supported many previous studies that attributed their positive results to using speech analysis technology as a main tool in phonetic teaching [14, 16, 19, 20]. However, the difference between previous studies and this study is that this study found no significant difference between the traditional and the modern teaching approaches, as both groups improved significantly after taking one of the training courses.

The improvement in participants’ pronunciation might due to the explicit information components and feedback. The role of speech visual displays alone might not have contributed to accurate production of sounds as the feedback from the instructor and Praat and repeating the sounds many times helped the participants improve as well. The results of this study provided support for literature in both traditional [13, 25] and modern teaching approaches [19, 20].

The two groups received the training conditions for four days (90 minutes/day), when each day was dedicated to one emphatic sound. Indeed, this amount of time spent in receiving explicit instruction was brief, but on par with the amount of time devoted to teaching phonetics explicitly in some previous research, which yielded significant and positive results [17, 23, 28].

4. CONCLUSION

This study has shed light on the possibilities of integrating technology, specifically speech analysis technique into Arabic pronunciation curriculum to enhance learners’ pronunciation of difficult L2 sounds. The contribution of the current study is to show that explicit instruction is a strong candidate for leading to pronunciation improvement, and that Arabic emphatics can be explained by presenting their visual representation forms.

The explicit information component, used in this study, need to be controlled in order to see better if using speech analysis in learning Arabic sounds would be significantly better than the traditional method. It appeared that explicit information and feedback were possible confounding variables that were not taken into consideration, eliminating the possibility to conclusively determine if the use of Praat actually helped Arabic learners or whether the explicit instruction and feedback improved learners’ perception and production. Future research could eliminate this limitation and control these variables.

It is hoped that this paper will benefit Arabic language instructors and researchers in embedding this modern tool in Arabic sounds teaching to enhance learners understanding which can be used inside or outside of classroom settings. Further research will include looking at the effect of using Praat with L2 learners of Arabic from different proficiency levels and from different language backgrounds to investigate the variations of perception and production abilities and to identify whether this tool is more beneficial to specific proficiency level or language group.
5. REFERENCES

QUESTIONING THE TEACHING OF “QUESTION INTONATION”: THE CASE OF CLASSROOM ELICITATIONS
CANTARUTTI, MARINA NOELIA

ISP DR JOAQUÍN V GONZÁLEZ / UNIVERSITY OF YORK
mnc510@york.ac.uk

ABSTRACT
In pronunciation textbooks for EFL teachers and learners, “question intonation” has been most widely associated to syntactic structure and attitudinal meanings. Even though these descriptions may refer to frequent realisations in real life, they appear to assume that all interrogatively-formatted utterances fulfil only informational functions and work similarly across different contexts. This reflective paper will make a critical review of the descriptions of “question intonation” in a selection of theoretical and practical materials employed in the Phonetics modules in three graduate Teacher Training programmes to contrast them to the findings of a short conversation/discourse-analytic study on teacher questions in classroom discourse, as representative of one of the speech genres EFL teachers are trained in. Finally, a case will be made for the need for further corpus-based descriptions of intonation that focus on social action in situated contexts of language use.

Keywords: intonation, questions, elicitations, teacher training, classroom discourse

1. INTRODUCTION

“There is no such thing as question intonation”, Cruttenden [7, p.59] categorically asserted. However, “question intonation” is presented as something apparently easily definable and predictable in many English Language Teaching instructional materials.

The simplifying assumptions found in these textbooks differ with the complex reality of question form and use, since questions “can be used for actions other than questioning, and questioning can be accomplished by linguistic forms other than questions” [26, p.34]. What is more, many of the accounts in these textbooks are based on de-contextualised and artificial examples, and arguably, many of the descriptions are introspective in nature, since there is no explicit reference to corpus work.

The current paper will present a review of “question intonation” in a selection of intonation manuals and activity books. The descriptions made in these materials will be problematised, and confronted with the results of a small-scale corpus study of teacher questions in one particular form of institutional talk: classroom discourse.

This study converges with previous discourse and interactional research on intonation [11, 18, 33, 29] and on intonation in questions [6, 31], that describe how prosodic choices both project, and are constrained by, the wider context of culture (i.e. genre) [10, 12], as well as by the specifics of the here-and-now emerging context of interaction. Therefore, this reflective paper will finally propose that descriptions of those intonation choices that have a certain regularity and systematicity should be presented to learners alongside the descriptions of the social and structural organisation of particular occasions of language use, with a focus on verbal and non-verbal social action [28].

2. QUESTIONS IN INTONATION MATERIALS

This section will review the description of “question intonation” in a selection of materials comprising the set bibliography in the Phonetics and Phonology modules at three Teacher Training Colleges in the city of Buenos Aires, Argentina. These materials are expected to inform teacher trainees’ choices of intonation in different speech situations, including “teacher talk”.

<table>
<thead>
<tr>
<th></th>
<th>Syntactic</th>
<th>Attitudinal</th>
<th>Discourse control / Info status</th>
<th>Linguistic action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baker [2]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil [4, 5]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Hewings [13]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>O’Connor and Arnold [22]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Wells [32]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Trench [25, 30]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Figure 1: Criteria associated to “question intonation” in pronunciation/intonation materials

2.1 Syntactic accounts of question intonation

Most of the materials scrutinised assume a one-to-one mapping between intonation and grammatical form. Falls are associated with wh-questions, and to the final item in a closed alternative question [22, 32, 2] and the use of non-falls is ascribed to the expression of attitudinal stance, making the question “more gentle, kindly encouraging” [32, p.43].

In [22, 32, 2], rises and fall-rises are said to belong in yes/no questions, and on the first items in closed
alternative questions, or on all the items in open alternative questions.

According to [32], it is the use of rises that marks declarative questions as such, though Brazil [4, 5] and Tench [29, 30] acknowledge the common use of falls in these. The inclusion of “declaratives” is the only hint in these materials so far that questions may not just be effected through interrogatively-formatted structures.

2.2 Attitudinal accounts of question intonation

Attitudinal approaches to tone include the association of intonational contours with displays of affect, emotion, and social solidarity or distance. This is a criterion some materials resort almost fully to [22] or combine with syntactic criteria [32], whereas others approach this in a functional way [4, 5, 29, 30].

Low falling tones on wh-questions are said to be indicative of “briskness”, “seriousness”, “urgency”, “disapproval” [22, 32], whereas high falling tones are “business-like”, “insistent”, “interested” [22, 32]. On yes/no questions, falls are used for “urgent discussion”, “scepticism”, “protest”, [22], as expressions of “surprise” or “pleasure” [13], or to establish forms of social divergence [4, 5].

Rises and fall-rises on wh-questions display “interest” and “friendliness” [22], but with a low head the speaker may seem “disapproving and resentful”. In wh- or yes/no phatic questions -enquiries about social wellbeing-, rising tones establish convergence and “insinuate togetherness” [4, p.79].

2.3 Functional accounts of question intonation

Functional descriptions of question intonation are far less common in the materials scrutinised and are associated to the traditions of either Discourse Intonation [4, 5, 26] or Systemic Functional Linguistics [29, 30]. Functional accounts could be broadly classified into two sub-criteria: information status and discourse control, and linguistic actions.

2.3.1 Questions and linguistic action

Of all the materials reviewed, only [29] makes an explicit, systematic connection between linguistic action and tone, especially in the realm of “suasion”, extending beyond information exchange. Falling tones emphasise the knowledge dominance of the speaker, and this is used in conducive questions -acting as covert opinions-, guesses, denials, prompts, lead-ins to stories or jokes, second-repeats, and second-attempt questions, and recommendations.

On the other hand, rises are found in invitations, requests, and offers, deferring authority to the addressee [29, 30]. For Brazil [4] these goods-and-services questions carry rises when asked “for the benefit of the hearer”, whereas when it is the speaker who “gains” from the answer, fall-rises are preferred.

The review of materials above has revealed the following underlying assumptions: a) questions are generally instantiated as interrogative forms; b) the grammar of the question may define its intonational contour; c) the intonation of a question may also be contingent on social, attitudinal or informational concerns; d) the intonation of questions is somehow predictable, beyond the context of use. And, based on the scrutiny of the examples provided in these textbooks: e) questions are generally produced as one intonation phrase (IP), except for alternative questions, and those with a “non-final” level tone [22] followed by a final rising or falling tone. In the sections that follow, the applicability of these descriptions will be evaluated in one particular institutional form of talk: classroom discourse.
3. QUESTIONS IN “TEACHER TALK”

The singularity of the structure of classroom interaction has been widely described [1, 8, 9, 18, 19, 25, 26, 27]. The most popular model establishes that the classroom exchange is typically organised around three moves: Initiation, Response and Follow-up (IRF) [26], and these generally occur in whole-class “recitations” of information testing and recalling [20].

In a corpus study, Alexander [1] found that recitations were generally brief and that “teachers moved from one child to another in rapid succession in order to maximise participation, or from one question to another in the interests of maintaining pace” (p.99). Since teachers often know the answers [14], students have to “match the questioner’s knowledge, or fall within those parameters” [17, p. 286], which is why these are known as test [20, 21], recitation [8, 9, 20] or display [14] questions. Two other types can be established in terms of a possible “prescriptiveness” of the answer: questions to which there is no closed answer are called referential [14], or authentic [20, 21]; and uptake questions, that incorporate a student’s previous answer as a way of “facilitating the negotiation of understandings” [21, p.146]. Conversation Analysis (CA) studies have observed that these question-types cannot be identified as such until the Follow-up move [18, 19], as it reveals the teacher’s orientation to the students’ answer, which in itself uncovers how the student interpreted the question.

A second CA-relevant classification centres around teacher control in speaker selection and turn-taking, and divides questions into individual nomination questions, invitations to bid, and invitations to reply [17]. Individual nomination questions are addressed to a particular student, whereas invitations to bid are questions requesting voluntary participation; invitations to reply, on the other hand, allow students to freely self-select to provide the answer straightforwardly. In the following section, these categories will be cross-matched to the descriptions of questions above.

4. METHODOLOGY

For this exploratory study, a short corpus (33 minutes) of Key Stage 2 British classroom interactions was scrutinised. The 90 teacher-led elicitation tokens found were collected from 7 whole-class recitations acting as revisions, or lead-ins. The lesson segments were selected from two YouTube channels: LessonsinObservation [15], and MediaMergeLtd [16]. Both companies claim to have obtained informed consent for online publication.

In-keeping with the qualitative methodologies privileged in DA and CA, questions were classified according to the types already discussed, and coding decisions made after turn-by-turn analyses of sequential positioning in the exchange, student orientation, and teacher feedback. Consistently with other studies of social action, non-interrogatively formatted items acting as questions were also included, so the term “elicitation” will be preferred.

Each question was transcribed, divided into IPs and its tone contours were specified according to those described in British school of intonation [22, 32]: fall, rise, fall-rise, rise-fall, level. Impressionistic observations were validated instrumentally with Praat [3]. The initial exploration of data revealed the presence of type-confated, multi-IP questions, which posed the need for further intonational detail, so the specification of IP numbers for each question, and the tone produced in each, were added.

Apart from qualitative analyses, basic inferential statistics were run in order to establish possible relations between terminal tone (i.e. tone in the last IP) and the different question types in the data.

5. RESULTS AND DISCUSSION

As expected, the data reveal several complexities that cannot be easily predicted by the above descriptions of question intonation. A selection of these issues will be addressed below.
A simple glance over the results might misleadingly confirm a straightforward relation between terminal tone and syntax, and with question type:selection, with a smaller likelihood of a connection between tone and question type:answer. However, statistics can be deceptive when these questions are seen in the sequential verbal and non-verbal context of their production, and in terms of the specifics of their formatting.

The fact that over half (57.8%) of the elicitations were produced as more than IP led to the problematisation of the role of terminal tone. Most questions were generally complex, made of an invitation to bid followed by a display question in the same syntactic structure, and these were also signalled by intonation: the invitations to bid had their own IP, with a fall-rise (58%), a level tone (25%), a fall (8.5%) or a rise (8.5%), and the upcoming display part of the elicitation generally carried a fall (92%), whether it was a yes/no, or a wh-question -including rear-loaded Wh-questions (Margutti, 2006 in [6]).

An examination of non-verbal behaviour also contributed to the questioning of the defining role of terminal tone. In some cases, the timing of anticipatory student bidding was found to be a marker of projectability, indicating that the function and content of the question were made relevant before the last IP was produced. This was also supported by teacher’s gaze direction as the question was being staged. Their monitoring of student bids and immediate nomination after the end of the question may render these final falls as markers of turn-transition, while the previous IPs with either falling or level tones may have been employed to give students time to process the question and bid. Thus, in this respect, tones appear to bear an organisational role over any grammatical or attitudinal association, with falling tones “advancing” [27] the lesson. This makes sense with what has been described for recitations, where the teacher’s control over the whole class and the preference for progressivity are foregrounded [1, 8, 9, 20, 21].

What is more, Tench’s [29, 30] reference to knowledge dominance and the notion of “conducive questions” (also covered in [6]) appear to be better suited to the discussion of display questions, as it cannot be argued that falling tones mark these as truly “finding out” [5], unless we consider display questions as elicitations whose information gap lies in the teacher’s need to know if the topic is clear, or easily recallable [8]. It is clear, nonetheless, that except for the “businesslike” nature of these questions in recitation, none of the attitudinal labels proposed in the above materials seem to characterise these classroom elicitations.

Rising tones were frequently found in the few (19%) individually-nominated yes/no uptake elicitations. Even though these are concerned with already-negotiated information, it is not the teacher’s own “making sure” function that is expressed, but they are presented as invitations to check, correct, or fine-tune the answers given, symbolically “deferring” to the learner’s knowledge to try again. Some of these were formatted as closed alternative questions, following the rising + falling patterns described.

The only clear syntax-tone correlation occurred in level-tone declaratives, which were in fact designedly incomplete utterances [DIU, 19], acting as “fill in the blanks” invitations to reply, answered in unison.

6. FINAL REMARKS

The use and formatting of elicitations is distinctive in “teacher talk”. Situated studies of questions in different episodes of everyday social action can, and have shed light into characteristic tonality and tone patterns that can only be made sense of if they are embedded on a particular occasion of talk, and not on “context-free notions of grammar, meaning and function”[28, p.150]. As ELT materials have done for the teaching of grammar, lexis and structure for different written genres, the role of intonation as an essential part of meaning-making in speech needs to be taught in close relation to in particular episodes of situated language use and speech styles[11], with a focus on social action.
6. REFERENCES


TECHNOLOGY- AIDED PRONUNCIATION TEACHING IN AN ESP/EAP COURSE

Ana M. Cendoya

ABSTRACT

Pronunciation has always been an important component of English language learning; however, little classroom time tends to be devoted to its teaching and learning in ESP / EAP courses. When non-native speakers deliver oral presentations in English, pronunciation is of paramount importance in order to successfully communicate the results of their research in international academic contexts.

This paper describes an ESP/EAP course for undergraduate students at the Engineering school, whose aim is to make effective oral presentations. In delivering the course, it has been observed that oral performances of their presentations pose serious problems as students encounter various difficulties regarding the perception and production of certain English sounds. The use of technology meant removing face-to-face class time constraints to deal with the teaching and learning of pronunciation more effectively. Furthermore, ICT has increased students´ knowledge of segmental phonetics and promoted intelligibility and academic appropriateness to their presentations.

Keywords: presentations, segmental phonetics, ESP/EAP, technology.

1. INTRODUCTION

1.1 Course Description

The experience described in this paper involves the delivery of an ESP/EAP course designed for Spanish speaking students at the Engineering School at Universidad Nacional de La Plata, Buenos Aires, Argentina. This is an optional subject and therefore students’ age ranges from 18 to 24. The number of students enrolled in the course varies between 45 and 60 and they are normally divided into three groups. As part of their research work they should make presentations describing their partial results, some of which are then presented at international academic events.

The course is aimed meeting the following students’ needs: specialized vocabulary of the field of Engineering, characteristic grammatical patterns of academic oral presentation such as passive voice, and the thematic structure of this oral genre. The course presents students with authentic and non-authentic texts which illustrate the type of language required to deliver oral presentations successfully and on which their productions will be based.

After performing skimming and scanning tasks of academic texts, students analyze and identify their lexical, grammatical and discourse features. Finally, students work with academic presentations – scripts and actual performances - in order to infer the thematic structure most of them present. Concurrently with the analysis of the thematic structure of presentations, the course attendees assess the differences in the way the language might impact the audience and the clarity of the message sent in each sample text.

Once the thematic structure of the genre is understood students start the script writing process of their future presentations. In this stage, the focus is on the careful choice of key vocabulary and grammar as well as the discourse features which best serve the pragmatic purpose of the presentation to be delivered. The following and last task then is the actual oral performance of the written scripts for assessment and feedback.

1.2 Identification of problems

Several authors have expressed their views in relation to the problems foreign language learners face when studying English and how important the teaching of pronunciation is in EFL courses. Celce-Murcia [3] states “Students will have oral communication problems no matter how excellent and extensive English grammar and vocabulary might be.”

Thus, there are sound reasons for teaching pronunciation and identifying the most important aspects of the segmental and supra segmental features which should be appropriately included in language courses. In the course here described students study English as a foreign language since they may have to
operate in international contexts and interact with native as well as non-native speakers. Jenkins [8] when referring to English as a lingua franca - a common language among people who do not share the same native language - considers that if intelligibility is to be achieved, the following characteristics of pronunciation should be included in English language courses:

• all the consonants, except /θ/ and /ð/
• initial consonant clusters
• vowel length distinctions
• the mid-central NURSE vowel
• nuclear stress

The attendees of this ESP/EAP course are mostly Spanish speakers; therefore, some of the difficulties identified during their first attempts at performing the presentations were the articulation of English initial clusters combining the voiceless alveolar fricative /s/ followed by voiceless plosives, nasals, semivowels or the lateral sound / p, t, k, m, n, l, w, j /, pronunciation of -ed endings and rhythm, stress and tonicity.

According to Coe [5], some of the major problems Spanish-speaking learners of English experience at the segmental level are “the difference between /b/ and /v/ as no difference is made; the production of /z/ /ʃ/ /ʒ/ and /dʒ/ as they do not exist, and voiceless plosives in initial position as they are not aspirated.”

With reference to word stress and rhythm, Finch [6] points out that “One of the differences between English and Spanish rhythm lies in the fact that Spanish vowel weakening is very slight compared with English.” The former has “a stress-timed rhythm because the accented syllables tend to occur at fairly regular intervals.” Whereas the latter has “a syllable-timed rhythm because it is the syllables, either accented or unaccented, which tend to occur at more or less regular intervals.”

All things considered, if the students attending the course were to succeed in presenting at international events, they needed specific training in order to overcome the difficulties above described.

2. METHODOLOGY
2.1 Listen and Repeat

The first strategy implemented in the course was reading aloud and repetition of key words, expressions and some complete utterances. Unfortunately, it failed to help students develop proprioceptive intelligence; defined by Underhill [9] as “internal kinesthetic awareness of the position and movement of the muscles and parts of the mouth which make the sounds.” In line with this, Fraser [5] states that “listen and repeat tends to generate limited successful results, it thereby creates a misguided image that pronunciation instruction can only be done if one possesses specialist knowledge of phonology and phonetics.” Furthermore, it may lead students to find the practice demotivating and frustrating at times.

Underhill [9] argues that “Having ‘good pronunciation’, and whether you are a native or non-native speaker teacher, are not the main issues. If you don’t know what you are doing in your mouth you are going to be restricted in the ways you can help a student.” In fact, it is also argued that “repeat after me or other types of practice exercises and discriminations do not necessarily develop proprioceptive intelligence [9].” Therefore, students cannot be freed from the oral grip of their L1 pronunciation habits.

2.2 Use of Technology

It was at this point in the course when technology shed some light on how to overcome the pronunciation problems identified. Teachers were enabled not only to foster students’ accurate pronunciation by means of marking aspects which needed improvement but also to give students feedback on their improved versions in due time. Furthermore, technology provided students with the autonomy to practise as much as necessary to improve their oral proficiency in and outside the classroom. In other words, students benefitted from the flexibility that technology added to their learning environment.

The first move towards the use of technology for the improvement of pronunciation was asking students to upload their recorded productions in a Moodle environment belonging to the course. Teachers listened to these oral texts and provided feedback in a new recording in which the correct pronunciation of words, phrases and complete utterances was illustrated. After receiving the teacher’s recorded feedback, students made an improved version of the original one which was again sent back to the teacher. This strategy then evolved in the design of an e-portfolio of oral texts defined by Cantarutti [2] as a collection which “illustrates students’ achievements and the processes involved in the production of these oral tasks.” Although this was quite an effective way to help students, it was too time consuming and still, students’ initial difficulties were not completely dealt with.
The next move then was to tackle the above mentioned difficulties – initial clusters, ed endings and rhythm, stress and tonicity – for which teachers suggested some web tools and apps. The IPA Phonetics App was the application chosen to help students raise awareness and develop proprioception, which was mainly used to improve pronunciation of initial clusters and –ed endings. This app provides a touch interface for exploring the International Phonetic Alphabet. A symbol in the chart shown is pressed and the user can hear an illustrative example and see ultrasound videos which accompany the sounds articulated.

While students watched the video in the app, they looked at their own mouths with the front camera of their cell phones in order to compare what they did and what the image showed when producing sounds such as /p/, /t/, /k/, /d/. Meanwhile, teachers monitored students and kept a record of the sounds students struggled to produce. In the final stage of the activity both teachers and students shared the results of their observations and drew some conclusions about the perception and production of the sounds rehearsed. With respect to developing proprioception Underhill [10, 11] contends that pronunciation teaching should “get out of the head and into the body”, that by making sounds “visible” segmental features might become less demanding for students to master. IPA Phonetics app is of invaluable help to achieve proprioception as it provides the user with both the sight and sound version of a given phoneme.

Later on, students identified the words containing the initial clusters or –ed endings which were mispronounced in their presentations and made a new recording of the utterances which featured the clusters mentioned before. The identification tasks of the mispronounced words or phrases were first performed in class. After that, students repeated the first recorded productions using Soundcloud.com, a web platform which enables users to record their voices and share them as a sequence of posts. Teachers added feedback on students’ improved audio files as written remarks or pop-up messages which students had revised taking account of the points raised by the teachers. Obviously, this process of revising and improving the recorded versions of presentations was done outside the classroom and it was only possible because it was aided by technology.

The last flaw to be tackled involved stress, rhythm and tonicity. Some students tended to reproduce their scripts assigning extra prominence, which led to a zero tone that revealed lack of interpretation during the delivery of the oral texts. Brazil [1] identifies five degrees of engagement (level 1 minimum engagement to level 5 maximum engagement) regarding the interaction established between the speaker/reader and the text, being levels 1 and 2 the ones characterized by this disengaged production.

To help students overcome this rather detached performance, they were presented with Soundcloud.com, a multi-modal corpus-like website which allows users to practise pronunciation through short excerpts of movies and TV series. A word or phrase is typed and the website produces short video clips that feature the item. The sequence of video clips can be paused so that users can repeat the phrase or utterance heard. By means of this website students became aware of the difference between English and Spanish in terms of stress placement, rhythm and tonicity. This comparison and the analysis of their improved recordings led them to avoid stressing grammatical words.

The use of Playphrase.me was assigned as homework. First, students had to find in the website examples of the phrases or part of utterances to be improved in their own recordings, and listen to them. Following this, they kept a record of the differences between their own versions and the ones found in the website. These findings were then shared and discussed in class. This exchange of information after having used the website helped students increase awareness of the mistakes made when dealing with stress, rhythm and tonicity.

The following are some examples from the students’ presentations in which I have used capital letters to show tonicity:

/HOW//DID//IT//ALL//START?/

Expected performance: HOW did it all START?

/EACH/ /ANDROID/ /VERSION/ /IS NAMED/ AFTER/DIFFERENT//CODES/

Expected performance: Each ANDROID VERSION is NAMED after different CODES

/HIGH VOLTAGE/ /IS/ /VERY/ /DANGEROUS/

Expected performance: HIGH VOLTAGE is very DANGEROUS

3. ANALYSIS AND DISCUSSION

This paper describes the experience carried out in an ESP/EAP course in which pronunciation teaching has been strengthened and pronunciation learning has been improved as a result of the introduction of technology - aided practices. The conclusions and
advantages hereby reported are of considerable importance although they are not conclusive results of academic research.

Overall it could be concluded that there was a significant improvement in students’ segmental production, which resulted in a higher rate of intelligibility. This was observed in the final recordings of the students’ presentations in their e-portfolios as well as in their actual performances delivered by the end of the course. All this was directly related to the amount of practice they undertook by means of the apps and website used.

One of the most significant improvements has taken place in the production of -ed endings, mainly the pronunciation of these endings after /n/ /l/ in participles such as designed, pulled. By the end of the course students still struggled to produce -ed participles following /t/ such as in looked.

As regards initial clusters starting with the voiceless alveolar fricative /s/ followed by voiceless plosives (p, t, k), they seemed to pose a more challenging problem possibly due to Spanish transfer.

At the beginning of the course stress, rhythm and tonicity represented an insurmountable difficulty; however, as students advanced they realized the negative effect of assigning extra prominence to an oral message. As a consequence, they avoided placing the focus, for example, on the verb to be in utterances containing passive structures, on prepositions or personal pronouns.

Finally, students reported higher levels of confidence when communicating in academic events. Interestingly, they described the practice of pronunciation mediated by technology as a very useful and motivating task.

CONCLUSION

Technology has been indisputably making a substantial contribution to pronunciation-teaching practice. Its inclusion in ESP/EAP classes have led to positive results but above all, it has offered an array of opportunities for the teaching and learning of pronunciation.

In this particular case, ICT has been significantly important as it has helped teachers and students to work in a rich environment which allowed the former to improve their teaching pronunciation practices and the latter to overcome their pronunciation problems. Furthermore, it boosted students’ motivation towards a practice which may become frustrating at times when it is not attached to significant and goal-oriented tasks.

The most important advantage is the fact that technology-aided pronunciation practice has promoted intelligibility and academic appropriateness to the presentations delivered by Spanish-speaking students of Engineering.

REFERENCES

A CORPUS-BASED ONLINE MANDARIN PRONUNCIATION LEARNING SYSTEM FOR CANTONESE LEARNERS: DEVELOPMENT, EVALUATION, AND IMPLEMENTATION

CHEN Hsueh Chu and HAN Qian Wen

The Education University of Hong Kong
hsuechu@eduhk.hk / veronicahan923@gmail.com

ABSTRACT

Hong Kong is a metropolitan city that serves as an international financial centre, giving rise to multi-lingual characteristics. In addition to Cantonese and English serving mostly as first and second languages, Hong Kong government encourages Mandarin as the third language under the ‘Biliteracy and Trilingualism’ language policy. This study developed a corpus-based online pronunciation learning system for Mandarin teachers, learners and researchers in order to better understand the major problems that Hong Kong Cantonese speakers encounter when learning Mandarin pronunciation. A pronunciation learning website was built based on the Mandarin spoken corpus; it contains roughly eight hours of recorded data of four spoken tasks completed by Hong Kong university students. A Mandarin tone-training program was developed using spoken data and learning resources from the learning system. The research findings will inform student learning and teaching practices and enhance teaching quality.

INTRODUCTION

The development of modern information technology has impacted corpus linguistics research and language learning. Most corpus linguists agree that the use of corpus linguistics could improve the efficiency of language learning and teaching because corpora can provide language learners and teachers with methods to improve pedagogical practice.

Corpora may encode language produced in spoken or written form. Spoken materials in the form of orthographic and/or phonemic transcriptions can be compiled into a spoken corpus that is searchable via computer [9]. It is believed that language in spoken form is a more effective basis for language learning than language in written form [5]. A number of spoken corpora have been constructed for various research and pedagogical purposes inside and outside of Hong Kong, such as British Academic Spoken English [3] and the PolyU Corpus of Spoken Chinese [10]. Nonetheless, it is difficult for users to focus on certain pronunciation features that speakers with different backgrounds produce; this is because speakers with diverse language backgrounds produce speech data with a great deal of variation [4].

When Hong Kong speakers learn Mandarin pronunciation, as indicated by Si [11], native Cantonese speakers tend to be confused by Mandarin sounds that are similar to certain Cantonese sounds, causing them to use sounds from Cantonese instead of from Mandarin. Fu and Lin [7] list several common mistakes that Hong Kong speakers may encounter when speaking Mandarin: 1) confusion of tone 4 of Mandarin; 2) confusion of tone 3 of Mandarin; 3) confusion of zh, ch, sh and z, c, s; 4) confusion of j, q, x; 5) confusion of nasal vowels and back nasal vowels; 6) omission of medial i and u; 7) confusion of h and k.

The ‘Biliteracy and Trilingualism’ language policy encourages Hong Kong citizens to learn Mandarin; however, resources for Mandarin learning are not adequate because education, media and political debates are not very supportive of Mandarin acquisition [8]. Most of the Mandarin corpus contains data in written form rather than in spoken form and is not accessible to Mandarin learners. Currently, there are no spoken corpora specifically designed for Hong Kong learners of
Mandarin to provide support for their Mandarin pronunciation learning. Therefore, this study developed a language learning system that integrates a Mandarin spoken corpus with speech data from Hong Kong learners of Mandarin.

CORPUS-BASED MANDARIN PRONUNCIATION LEARNING SYSTEM

The corpus-based Mandarin pronunciation learning system [13] was developed to provide both authentic spoken data from Hong Kong learners and phonological features by which both Mandarin learners and teachers could access real speech data and scientific description. The system consists of three interrelated components apart from the core corpus: 1) online pronunciation practices specifically designed for Hong Kong learners that include interactive quizzes, songs, jokes and tongue twisters; 2) useful resources for Mandarin pronunciation learning and teaching, including an introduction to Chinese phonology, recommended learning websites, online videos and online dictionaries; and 3) Mandarin tone learning exercises for learners to practice Mandarin pronunciation with hands-on acoustic analysis techniques.

The free online software ‘Praat’, developed by Boersma and Weenink [2], was introduced to foster Mandarin pronunciation learning and phonetic research as an essential component of the learning system. A Praat manual for beginners derived from a previous study by Chen and Wang [4] of a corpus-based English pronunciation learning system helps users to access and navigate the learning system. The current learning system provides software demonstrations and additional acoustic analyses of Mandarin on the ‘Tone Learning by Praat’ page. Simplified instructions and a demonstration video are included to help users understand the basics of conducting acoustic analyses. Sample analyses of monosyllabic words, multisyllabic words and sentences produced by native Mandarin speakers from Beijing and Taiwan are given. Users can refer to these sample analyses and download native speakers’ and learners’ pronunciation of Mandarin words and sentences for their own acoustic analysis practice in order to improve their Mandarin tone pronunciation.

Spoken corpus of Hong Kong learners

The Mandarin spoken corpus [14] created by Hong Kong learners comprises the core of the learning system. It provides 40 sets of high-quality recordings to aid researchers in phonetic and acoustic analysis; it features annotations of the segmental and suprasegmental features that support learners’ search for specific phonological features and provides access to the corresponding audio files with annotated notes. The main goal of the corpus is to raise pronunciation awareness of Mandarin learners in Hong Kong and overseas, alerting them to their specific pronunciation problems and enhancing their active engagement in the Mandarin learning process.

The spoken corpus contains eight hours of recorded data of Mandarin spoken tasks (reading of monosyllabic words, multisyllabic words, passages and free speech) from 40 university students in Hong Kong with detailed phonological annotations of the recordings mainly focused on segmental and suprasegmental features. Users can browse the recordings by speakers’ demographic information (e.g. gender, age and major) and search the recordings by the pronunciation features they are interested in.

The speech data in the spoken corpus is organised based on four Mandarin spoken tasks: Task 1: Reading of monosyllabic words; Task 2: Reading of multisyllabic words; Task 3: Reading of passages; Task 4: Free speech. The design of Task 1 is based on the rationale that the task should cover all tones with even distribution and avoid polyphones and overlap with words in other tasks. Task 2 includes confusing clusters (e.g. Tonè+Tonè, Toné+Tonè, etc.), confusing vowels (e.g. nasal vs back nasal vowels), difficult consonants (e.g.
retroflex consonants ‘zh, ch, sh’, nasal consonant ‘n’, etc.) and tone sandhi (e.g. tone sandhi of ˇTone, ‘yi’, ‘bu’ and use of the neutral tone). Task 3 is designed to collect comprehensive speech data that reveal speakers’ pronunciation of vowels, consonants, tones, tone sandhi, etc. Task 4 collects the authentic speech data of Mandarin learners for research concerning discourse, phonology, grammar, vocabulary, etc.

**EVALUATION OF THE LEARNING SYSTEM: A MANDARIN TONE TRAINING PROGRAM**

The speech learning model suggests that the learnability of sounds in a second language (L2) depends on their similarity to their counterparts in the first language (L1), meaning it is easier for learners to master new sounds in L2 than similar sounds, which may confuse them [6]. Like Mandarin tones, Cantonese tones include level, rising and falling tones; therefore, Cantonese learners of Mandarin tones tend to rely on pre-existing L1 Cantonese tones [1]. In a study that So [12] conducted on native Cantonese and Japanese speakers’ perceptions of Mandarin tones, Cantonese speakers tended to have more difficulty distinguishing between Mandarin tones 1 and 4 and tones 2 and 3 than speakers of Japanese, which is a non-tonal language.

With the resources provided by the Mandarin pronunciation learning system and Praat, this study developed a Mandarin tone training program to evaluate the effectiveness of a tone training scheme for learners. The Mandarin spoken corpus provides users with a platform to learn from the pronunciation errors made by Cantonese-speaking learners of Mandarin. To learn Mandarin tones, learners can use the search function to look for incomplete tones, tone changes and tone sandhi features from the suprasegmental category. They can also learn Mandarin tones by viewing the annotations in each individual reading task by filtering the tone errors they are interested in from the feature list.

**Method**

Five university students who were native Cantonese speakers participated in the program. The program consisted of four parts: a pre-test, tone visualisation training, a post-test and tone rating. In the pre-test, the participants began by recording themselves reading a list of words in a computer laboratory. They then listened to a recording of the words from the reading task in random order and wrote down the tones they heard. In the tone training phase, the participants completed five on-campus training sessions and two online self-learning sessions over one and a half months using the acoustic phonetic software Praat and authentic speech data from the Mandarin spoken corpus and website. In the post-test, the participants recorded themselves reading the same word list as they did in the pre-test. The participants’ tone production in the pre-test and the post-test was assessed aurally. Three native speakers of Mandarin rated the participants’ tone production from both the pre-test and the post-test recordings. A follow-up interview was conducted with each participant to evaluate the effectiveness of the online platform and the training program.

**Results**

Table 1 presents the means of the correct rates for each Mandarin tone in the listening and reading tasks in the pre-test and the post-test. Compared to the results in the pre-test, the correct rates of all tones in both the listening and the reading tasks increased, except for the correct rate of pronunciation of tone 2 in the listening task, which remained exactly the same. Improvement in the correct rate of Mandarin tones was more obvious in the reading task than in the listening task. Correct rates of tone 3, tone 4 and the neutral tone in the reading task improved by 3.33%, 6.67% and 10%, respectively.
Table 1. The correct rate of each tone in the pre-test and post-test

<table>
<thead>
<tr>
<th>Tone</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Listen</td>
<td>Read</td>
</tr>
<tr>
<td>1</td>
<td>91.67%</td>
<td>85.83%</td>
</tr>
<tr>
<td>2</td>
<td>89.17%</td>
<td>85.00%</td>
</tr>
<tr>
<td>3</td>
<td>87.50%</td>
<td>79.17%</td>
</tr>
<tr>
<td>4</td>
<td>80.83%</td>
<td>88.33%</td>
</tr>
<tr>
<td>0</td>
<td>92.50%</td>
<td>80.00%</td>
</tr>
<tr>
<td>Mean</td>
<td>88.33%</td>
<td>83.67%</td>
</tr>
</tbody>
</table>

In the speakers’ reflections on the tone training program and use of the Mandarin pronunciation learning system, all five speakers pointed out that the authentic speech data from the Mandarin spoken corpus integrated into the training program with Praat was closely connected to their own Mandarin pronunciation learning experiences. Four of the speakers were excited to experience a new Mandarin pronunciation learning approach with the help of the resources from the learning system. Two speakers enjoyed the flexibility provided by the online self-learning exercises in the tone training program, which helped them to fully explore the Mandarin spoken corpus and perform simple acoustic analyses of other Hong Kong learners of Mandarin who shared similar pronunciation features.

CONCLUSION

The Mandarin pronunciation learning system was built to identify learners’ recurrent difficulties in accurately and appropriately using Mandarin segmental and suprasegmental features and to suggest possible solutions to reduce such difficulties. The results of this study reveal that learning resources, pronunciation exercises, the Praat user guide and authentic speech data in the corpus can be flexibly integrated into Mandarin pronunciation learning, teaching and research.

REFERENCES


REVIEW OF THE LINGUA FRANCA CORE FOR ENGLISH PRONUNCIATION TEACHING IN KOREA

Hyunsong Chung
Korea National University of Education
hchung@knue.ac.kr

ABSTRACT

This paper reviews the LFC (lingua franca core), which was suggested for English pronunciation in ELF (English as a lingua franca) contexts, and examines its implications in pronunciation teaching in Korea. According to the LFC, the intelligibility of the English pronunciation of L2 English speakers and English learners is more important than pronunciation accuracy in English communication. The paradigm of the LFC has been widely accepted in Korean classroom situations by teachers of English, so the systematic teaching of English pronunciation for accuracy has been rarely emphasized. This paper argues that the LFC of English is frequently misinterpreted in Korea, while some of the features in the LFC themselves are also misleading. This paper tries to redefine the LFC and provide suggestions for teaching pronunciation in Korea.

Keywords: Lingua franca core, LFC, pronunciation, intelligibility, Korean

1. INTRODUCTION

According to [6]’s LFC, intelligibility between NNSs (non-native speakers of English) has more priority over than between NNSs and NSs (native speakers of English); therefore, [6] emphasizes intelligibility over accuracy. Furthermore, [6] notes that most segmental features are important in the LFC, but [6] disregards the importance of suprasegmental features except for nuclear stress. [6]’s suggestions for the LFC in ELF contexts are summarized in Tables 1 & 2.

[6] argues that English learners’ regional accents due to suprasegmental features must be allowed because errors in suprasegmental features do not hinder the intelligibility of international communication. These features are usually difficult to teach and learn in order to enhance communication among NNSs, [6] argues. [6] even insists that certain features of English as a Native Language (ENL) pronunciation are detrimental to intelligibility in ELF communication. Typical examples of these detrimental elements are the use of weak forms and other features of connected speech, such as deletion and assimilation, which are usually connected with the stress-timing property of the English language.

This leads to the conclusion that the nuclear stress within an intonational phrase (IP) is the only important suprasegmental element in ELF communication.

Table 1: ELF pronunciation targets: core features (adapted from [7]: 23).

<table>
<thead>
<tr>
<th>ELF target</th>
<th>Lingua Franca Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonantal inventory</td>
<td>- all sounds except /θ/, /ð/ but approximations of all others acceptable</td>
</tr>
<tr>
<td>- rhotic /r/ only</td>
<td></td>
</tr>
<tr>
<td>- intervocalic [t] only</td>
<td></td>
</tr>
<tr>
<td>Phonetic requirements</td>
<td>- aspiration after /p/, /t/, /k/</td>
</tr>
<tr>
<td>- appropriate vowel length before fortis/lenis consonants</td>
<td></td>
</tr>
<tr>
<td>Consonant clusters</td>
<td>- word initially, word medially</td>
</tr>
<tr>
<td>Vowel quantity</td>
<td>- long-short contrast</td>
</tr>
<tr>
<td>Tonic (nuclear) stress</td>
<td>- critical</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ELF target</th>
<th>Lingua Franca Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel quality</td>
<td>- L2 (consistent) regional qualities</td>
</tr>
<tr>
<td>Weak forms</td>
<td>- unhelpful to intelligibility</td>
</tr>
<tr>
<td>Features of connected speech</td>
<td>- inconsequential and may be unhelpful</td>
</tr>
<tr>
<td>Stress-timed rhythm</td>
<td>- unnecessary</td>
</tr>
<tr>
<td>Word stress</td>
<td>- can reduce flexibility/unteachable</td>
</tr>
<tr>
<td>Pitch movement</td>
<td>- unnecessary/unteachable</td>
</tr>
</tbody>
</table>

[7] also suggests a pronunciation model of interlanguage talk (ILT) and same-L1 talk where accommodation skills and phonological convergence take place. According to this model, in ILT where learners of English with different language backgrounds engage in communication, the interlocutors try to accommodate their pronunciation in accordance with the other interlocutors’ pronunciation and phonological convergence takes place to enhance intelligibility and communication. Conversely, in the same-L1 talk, though phonological
convergence happens in order to improve communicative efficiency and solidarity, the speakers do not need to change their own phonological system because they are familiar with their own accent. Even within their own phonological system, they would have no problem communicating with each other, thus pronunciation is further fossilized.

[9] explores some of the concerns that surround the adoption of ELF and describes a number of potential benefits of an ELF approach to pronunciation. According to [9], the LFC could provide learners with a lighter workload and increased levels of achievability which in turn leads to increased learner motivation. It also helps learners maintain their first-language identity when speaking English. [9] also suggests various techniques and materials for teaching the LFC.

Although the LFC is an attractive approach in teaching pronunciation in that it focuses more on intelligibility than accuracy and that it pursues learnability, teachability, learner motivation and identity, its application in the Korean context has practical problems. Thus, it is necessary to redefine the LFC and its teaching techniques in the Korean context.

2. ISSUES OF THE LFC IN KOREA

Some features of the LFC are quite misleading. [6], for example, deliberately excludes English NSs from ELF interlocutors. Because there are approximately 400 million native speakers of English [4], they are still important participants in English communication as NS/NNS interaction still likely transpires. Even though there are more NNSs than NSs, Korean learners of English, specifically, are more likely to have communications with NSs than with NNSs outside their L1. Even though [6] argues that NSs must endeavor to understand the LFC of NNSs, not vice versa, the fact cannot be denied that the pronunciation of NSs has priority over that of NNSs in certain areas. Therefore, the LFC for communication among only NNSs could lead to serious communication breakdowns when NS-NNS communication occurs. The following example shows a momentary communication breakdown in interlanguage talk between a Korean speaker of English (KS01) and a native English speaker (ES01) due to the non-canonical pronunciation of the Korean speaker. The recording of the conversation was conducted in London, U.K.

(KS01_ES01, 6:20-40)
ES01: Are you, Are you studying here in this, in this country?
KS01: Yes .. uhm ..., I just studying English

In this conversation, KS01 pronounces ‘Leicester Square’ as [ɾɛstə skeə], where /l/ is replaced by [r] and /kw/ cluster is simplified to [k]. Even though ‘Leicester Square’ is not an unfamiliar place to ES01, she has no idea what KS01 intends to say until she gets a clue from ‘Charing Cross Road,’ which is located just next to ‘Leicester Square’ in London. For many Korean speakers of English, [r] could be regarded as an approximation of /l/. But a slight deviation from the canonical /l/ seems to hinder the communication in this example.

In another conversation between KS06 and ES03, K06 replaces /θ/ with [s] and omits [w] in the /gw/ cluster in the initial stage of the conversation. The affricate /dʒ/ is also replaced by the fricative [z].

(KS06_ES03, 6:36-7:00)
KS06: So you can speak three [sri], three [sri] language [ɛŋgiziz]?
ES03: Uh, speak ...?
KS06: Three languages [sri ɛŋgiziz]? Three languages [sri ɛŋgiziz], three [ɔri] language [ɛŋgiz].
ES03: Um ... Three ...
KS06: Languages [ɪŋgwiʤiz].
ES03: Yeah, just about it. Just about, I can speak three languages and reading and writing.

(adapted from [2]: 251)

The replacement of /θ/ and /dʒ/ to the fricatives is commonly observed in Korean speakers’ utterances in English. The fricative /θ/ is one of the non-core features in the LFC and the replaced [s] is an approximation of the segment. However, ES03 has difficulty in understanding the simple phrase ‘three languages’ due to the minor deviant pronunciation. These examples clearly show that even the non-core features of the LFC and the approximation of some consonants could lead to unintelligible pronunciation. [6] also argues that once the contrast is maintained between long and short vowels and the vowel quality is used consistently, the non-canonical forms of vowel sounds would be sufficient in ELF communication. This argument has been misinterpreted by many English teachers in Korea to
mean that any vowel quality would work in English communication, ignoring the fact that the contrast of the vowel quality should be consistent. The tense and lax vowel distinction is important even in the LFC. It is difficult for Korean learners of English to use consistent vowel quality. In order to handle the contrast of English vowels efficiently, accuracy-based pronunciation teaching should be enforced. Otherwise, it might be difficult for Korean speakers of English to acquire intelligible pronunciation of English vowels.

As shown above, most vowels and consonants in the LFC are more ‘native’ or ‘native-like’ pronunciation. Therefore, at least in regard to segmental features, the LFC differs little from ‘native’ English pronunciation. Even though it is unnecessary to acquire Received Pronunciation (RP) or General American (GA), learners must strive towards ‘native-like’ pronunciation even in the ELF context.

[6] further implies that reduced vowels, phonological changes in connected speech, stress-based English rhythm, and word stress might be unimportant phonological features for NNSs. It is argued by [6] that word stress should be removed from the LFC because it is difficult to teach and learn. However, too much deviation of such prosodic features would make it difficult for NSs or even NNSs to understand NNSs’ pronunciation. For example, misplaced stress on ‘normally (norMALLy)’ may be misheard as ‘no money’; misplaced stress on ‘written (wriTTEN)’ may be misheard as ‘retain’; and misplaced stress on ‘secondary (seCONdary)’ may be misheard as ‘country’ [1]. This happens because when a stress moves to another syllable it involves vowel quality change as well as a stress shift.

[6]’s disregard of suprasegmental features in the LFC is also criticized by [5]. [5] argues that just several English stress rules could explain the stress patterns of more than 85% of multi-syllabic English words. In addition, aspiration, vowel lengthening, and nuclear stress which are included in the LFC cannot be explained without referring to word stress. [5] also suggests that reduced vowels and phonological changes in connected speech are important because they contribute to easing speakers’ effort and enhancing ‘non-native’ speakers’ fluency. Improving ‘non-native’ speakers’ pronunciation of suprasegmental features leads to improved intelligibility by ‘native’ listeners [8].

An additional problem relating to [6]’s LFC is the misinterpretation of ILT by English teachers in Korea. In Korea, [6]’s LFC and ILT are misinterpreted to mean that once intelligible communication is guaranteed among ‘non-native’ homogeneous interlocutors, pronunciation is relatively insignificant. However, this could be a fatal flaw in English learning. What is intelligible pronunciation between Korean learners might not be quite understood by other NNSs or NSs. Because communication usually takes place in the same-L1 talk in Korean classrooms, the phonological convergence is likely to lead to the phonological fossilization than to the LFC. Additionally, the intelligibility of Korean accent in international communication has rarely been tested, and is paucity, at best, of empirical and thorough studies about how much of Korean students’ English pronunciation is intelligible. The precondition of [6]’s ILT model is that of the multi-lingual setting and thus the accommodation skills and phonological convergence suggested by [6] do not mean maintaining Korean accent, but rather learning a new phonological system that would be much closer to ‘native’ English pronunciation than to Korean accented English. Since it is difficult to apply the ILT model in the Korean context, a different effort to find an alternative pronunciation instruction model is needed.

One of the solutions to teach the LFC to monolingual groups is via student recording as suggested by [9]. In the recording, students can record a text and give feedback to their classmates focusing their attention on a small number of pronunciation points from the LFC and then practicing these in other activities. [9] also suggests an exposure to a range of ELF accents using multi-media resources. The critical problem of these approaches is that there is no interaction with NNSs. Peer feedback and multi-media resources do not provide the students with the ILT environment. It improves learners’ receptive skills more than productive skills.

We cannot simply ignore the socio-linguistic aspect of English pronunciation. Intelligible but accented pronunciation is not necessarily what learners of English want to learn. There are many occasions in the job market that biased preference goes toward less regionally-accented English speakers. That’s why many learners of English try to acquire ‘standard’ or ‘native-like’ English pronunciation. This is the same with English teachers in Korea, and is ironic in that they rarely teach their pupils ‘standard’ or ‘native-like’ pronunciation in classrooms. Students should be exposed to a ‘native-like’ reference model in English classrooms. Even in [7]’s survey about the descriptions of English accents by expanding-circle and native English speakers, the Korean English accent was described as one of the most noticeable variants. The negative comments were much higher than the positive comments compared to other accents. It was described as ‘difficult to understand, strange, harsh, nasal and quarrel like.’

So accuracy as well as intelligibility are important
factors in learning English pronunciation. English learners’ accented pronunciation usually has more correlations with suprasegmental features such as intonation and speaking rate than with segmental features [8]. Therefore, more emphasis should be put on suprasegmental features to enhance the accuracy and the intelligibility of English pronunciation.

3. CONCLUSION

[6]’s LFC should be modified and appropriately interpreted in the Korean context. Contrary to [6]’s intention, [6]’s suggestion of the LFC is much closer to ‘native’ English speakers’ pronunciation than to the Korean accent of English at least in segmental features. Suprasegmental features in English, which were disregarded by [6], are also as important as segmental features to improve intelligibility. As suggested by [5], many suprasegmental features are learnable and teachable. Socio-linguistically, ‘native’ or ‘native-like’ English pronunciation is still generally preferred to accented English pronunciation, and thus in this paper it is concluded that English pronunciation syllabi in Korea still need to provide English learners with ‘native-like’ pronunciation including suprasegmental features as a reference model.

The LFC could be allowed if it does not hinder communication and if intelligible pronunciation has priority over accuracy. However, intelligibility and accuracy are not exclusive to each other as they may be addressed simultaneously. In terms of listening, it is necessary to develop materials which contain ‘non-native’ norms. The materials should be more authentic than before so that Korean learners of English may successfully cope with interlanguage talk. However, in terms of production, it still needs to adhere to a ‘native-like’ model based on the argument given above. In English classrooms in Korea, pronunciation teaching is conducted contrary to this concept. A ‘native-like’ model and norm is given to the students to improve their listening ability, while fossilization is allowed in production. Regarding L2 learners’ learning of English rhythm, [3] suggests that L2 learners who have syllable-based rhythm would become competent in both syllable-based and stress-based speech, continuing to use syllable-based speech for local communication, as a signal of national identity, and switching to stress-based speech for international communication, as a means of ensuring intelligibility. [3]’s suggestion could be applied to teaching other pronunciation features. Learners may not reach ‘native-like’ pronunciation proficiency and the norm of English pronunciation would be the LFC. However, the reference model or input to English learners in classrooms must move toward ‘native’ or ‘native-like’ pronunciation to guarantee intelligibility in international communication in any context. Then, learners can switch between ‘native-like’ norms and LFC norms when either norm is required.

4. REFERENCES

HOW IMITATION CAN HELP THE ACQUISITION OF L2 PRONUNCIATION

Eva Estebas-Vilaplana

Universidad Nacional de Educación a Distancia
eestebas@flog.uned.es

ABSTRACT

The aim of this paper is to examine the production of English stress-timed rhythm and the weak vowel schwa by Spanish students of English when they speak L2 English and L1 Spanish with an imitated English accent. The hypothesis is that Spanish students of English have a better command of English rhythm and weak vowels in imitated speech than in L2 speech. The productions of four Spanish students of English in three reading conditions (L1 Spanish, imitated English and L2 English) were analysed and compared to those of English speakers. The results showed that the Spanish speakers’ productions of English rhythm and the schwa are closer to “real” English in the imitation scenario than in the L2 speech. These findings support the idea that a top-down approach to phonetics teaching (from rhythm to segments) may be more beneficial for languages with a syllable-timed rhythmic pattern, such as Spanish.

Keywords: rhythm, stress, weak vowels, L2 pronunciation, L2 imitation.

1. INTRODUCTION

Over the past few years phonetic training has become more and more important in L2 English classes in Spain. Given the significant differences between the English and the Spanish phonetic systems, L2 English teachers have realised that if L1 Spanish speakers want to achieve a successful communication in English, they should be acquainted with a few phonetic notions ([8]). The incorporation of some pronunciation guidelines and orientations in L2 English classes has taken place both in adult ([7]) and in childhood education ([5]). In both scenarios, phonetic training has been focused on segmental issues, namely, on the identification of the most problematic sounds along with their corresponding phonetic symbols. In most cases, the methodology used to present the phonetic contrasts has been based on minimal pair work and on listening activities.

Even though the presence of phonetic tuition in L2 English classes in Spain has represented a strong step forward to enhance the intelligibility of Spanish students of English, there is still a long way to go before achieving a fluent and effective communication. L2 English students have indeed become more aware of the differences between English and Spanish pronunciation in terms of vowels and consonants and they know how to identify and produce English sounds in specific words. However, when they move on to connected speech and they do not specifically focus on the pronunciation of a particular word, their phonetic knowledge seems to disappear since they no longer produce the expected vowel and consonant contrasts they used to produce in isolated words. Given this observation, the necessity of changing the priorities in phonetic teaching begins to take shape. If the main problems arise in connected speech, maybe phonetic tuition should not start at the sound-word level but at a higher level, such as the sentence level. This means that instead of first teaching vowel and consonant contrasts in an L2 English class, it would be interesting to see how beneficial for L2 students would be to first become acquainted with supra-segmental features.

The idea of using a top-down approach in phonetics teaching has already been pointed out in some works, such as [1]. This investigation proposes to use songs and poetry, that is, texts with a rhythmic pattern to start learning phonetics and thus move from the rhythmic structure of sentences to the segmental structure. Similarly, [12] suggests that an appropriate stress and rhythmic pattern is more important for intelligibility than the correct pronunciation of isolated segments and, in fact, stress and rhythm determine the pronunciation of segments in English.

The aim of this paper is to provide some more evidence to the idea that rhythm can be a good point of departure to teach English phonetics to foreign students, especially when the rhythmic pattern of their L1 differs from that of English. This is the case of Spanish and English which have a syllable-timed rhythm and a stress-timed rhythm respectively ([4]). In order to prove how relevant rhythm is for phonetics teaching, we will compare the speech of four Spanish students of English speaking L2 English, and speaking L1 Spanish with an English accent, that is, in imitated English-accented Spanish.

The idea of comparing these two types of speech arose from the observation that a lot of people are able to imitate an L2 accent when they speak their L1. Thus, they are able to perceive and reproduce some of the phonetic features of the L2 in their imitation without having had any specific phonetic training. However, when they speak the real L2 some of the
phonetic features they are able to imitate do not emerge. Obviously, the lack of fluency and the lack of vocabulary may also play an important role. The clear mismatch in the pronunciation of an imitated non-native accent and the real L2 accent is an area that will be investigated in this paper.

Thus, this paper examines the differences in the phonetic features used in an imitated English-accented Spanish and in the real L2 English. Particular attention will be paid to the production of rhythmic patterns and the weak vowel schwa since one of the features that “good” Spanish imitators of English pronunciation seem to recognize is the importance of stresses in English. According to this observation, this paper explores two hypotheses: 1) Spanish students of English are better at reproducing English rhythm and weak vowels in imitated English-accented Spanish than in L2 English, and 2) rhythm is a salient feature in imitations with a positive effect on the production of weak vowels. The results of this investigation will shed some light on whether a top-down approach to English phonetics teaching can help to improve the English pronunciation of Spanish speakers.

2. THE EXPERIMENT

2.1 The materials

In order to analyse the role of rhythm and its effects on vowel quality both in imitated English-accented Spanish and in real L2 English, two phonetically balanced texts were used, one in Spanish (as in [3]) and one in English (as in [6]). The two texts were divided according to their feet. A foot is a unit of rhythm which begins with a stressed syllable and is followed by zero or more unstressed syllables (see [10]). In the two texts, the number of unstressed syllables per foot ranged from 1 to 6. For this study, only feet with 2, 3 or 4 unstressed syllables were taken into account, since feet with only 1 unstressed syllable or feet with more than 5 were less frequent. The data was elicited by means of a reading activity so that students only focused on the pronunciation and not on the vocabulary or the grammar.

2.2 The informants

Four Spanish students of English with an A2-B1 English level according to the Common European Framework of Reference for Languages and two English speakers performed the experiment. The Spanish speakers were two male and two female students from the north of Spain aged between 41 and 49. The two English subjects, a male and a female speaker from the south of England aged 45 and 35 respectively, were included as a group of control to compare their productions with those of the Spanish speakers. The Spanish informants recorded the texts in a soundproof room at the Universitat de Girona and the English subjects at the Universidad Nacional de Educación a Distancia in Madrid. The Spanish students read three texts: 1) the Spanish text in Spanish, 2) the Spanish text with an imitated English accent and 3) the English text in English. No specific instructions were given regarding imitation so the speakers had to rely on their stored knowledge on English pronunciation. The English informants only read the English sample. The Spanish text read in L1 Spanish and the English text read in L1 English were used as points of reference for the analysis of the rhythmic pattern and the weak vowel productions in imitated English-accented Spanish and in L2 English.

2.3 Data analysis

For all the productions, two parameters were examined: 1) the rhythmic pattern and 2) the quality of the vowels in the unstressed syllables. For the analysis of the rhythmic pattern, each text was divided into its feet according to the number unstressed syllables (from 2 to 4). For each foot the duration of the syllables between stresses was measured by means of the Praat program ([2]). The cases of anacrusis and the cases of unstressed syllables before a prosodic break were not taken into account. For example, in the following Spanish sentence (“The jeweller Federico Banero has been condemned by the Santander Court”) only the duration of the syllables in bold was measured. Stressed syllables are underlined and the number of unstressed syllables for each foot is indicated below.

(1) El joyero Federico Banero / 3 ha sido condenado por la audiencia de Santander / 3 4

The measures were normalised so that they could be compared irrespective of the speaker’s speech tempo. Normalisation was done by dividing the duration of the unstressed syllables by the duration of the whole sentence.

For each foot, the vowel quality of the unstressed syllables was assessed by means of an auditory analysis performed by three specialists in pronunciation, two Spanish phonetic teachers of English and one native English teacher. Only the weak vowel schwa (/a/) was taken into consideration since the production of this sound is more problematic for Spanish students of English than the production of the other English weak vowels. A separated audio file was created for each foot, so that the assessors could concentrate on a short speech chunk. They could hear the file as many times as
necessary and they were asked to mark the schwa productions on a list containing the different feet.

3. RESULTS

3.1 The rhythmic pattern

Table 1 presents the mean duration in seconds of the unstressed syllables in feet with 2, 3 and 4 unstressed syllables for the four Spanish students (S1-S4) in the three reading conditions: Spanish text in Spanish, Spanish text with an imitated English accent and English text in L2 English. The results for the English speakers are presented in Table 2 for the English text.

For all speakers (both Spanish and English) and for all reading conditions, the results show that the duration of the unstressed syllables between stresses increases as the number of the syllables increases. These results are consistent with the findings of other studies ([9], [11]) which showed that the distinction between syllable-timed languages and stress-timed languages is not so clear-cut and that the regular occurrence of stresses in languages with a stress-timed rhythm, as English, is not “perfect”.

Table 1: Mean duration (in seconds) for 2, 3 and 4 consecutive unstressed syllables in the English text produced by four Spanish speakers (S1-S4).

<table>
<thead>
<tr>
<th>Sp.</th>
<th>Spanish text in Spanish</th>
<th>Spanish text in imitated English</th>
<th>English text in L2 English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>S1</td>
<td>2</td>
<td>.3</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>S2</td>
<td>2</td>
<td>.3</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>S3</td>
<td>2</td>
<td>.3</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>S4</td>
<td>2</td>
<td>.3</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2: Mean duration (in seconds) for 2, 3 and 4 consecutive unstressed syllables in the English text produced by two English speakers (E1-E2).

<table>
<thead>
<tr>
<th>Sp.</th>
<th>English text in L1 English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>E1</td>
<td>.25</td>
</tr>
<tr>
<td>E2</td>
<td>.21</td>
</tr>
</tbody>
</table>

Despite the relationship between the number of syllables and their duration in the two languages, more relevant results are observed if we compare the duration of the time interval between 2 and 3 unstressed syllables and 3 and 4 unstressed syllables. As presented in Table 3, the mean duration interval between 2-3 and 3-4 unstressed syllables is larger in all the Spanish productions than in the English ones. These results show that Spanish speakers take more time than English speakers to produce the equivalent number of unstressed syllables between stresses, supporting the tendency of a syllable-timed rhythm in Spanish as opposed to the stress-timed English rhythm.

The next question, however, is to analyse what Spanish speakers do when they speak L1 Spanish with a imitated English accent and when they speak L2 English. Table 4 shows the overall and individual mean duration interval between 2-3 and 3-4 unstressed syllables in imitated English-accented Spanish and in L2 English produced by Spanish speakers.

Table 3: Overall and individual mean duration interval (in seconds) between 2-3 and 3-4 unstressed syllables in Spanish (S1-S4) and in English (E1-E2).

<table>
<thead>
<tr>
<th>Sp.</th>
<th>Spanish text in Spanish</th>
<th>English text in English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-3</td>
<td>3-4</td>
</tr>
<tr>
<td>S1</td>
<td>.14</td>
<td>.27</td>
</tr>
<tr>
<td>S2</td>
<td>.12</td>
<td>.28</td>
</tr>
<tr>
<td>S3</td>
<td>.12</td>
<td>.24</td>
</tr>
<tr>
<td>S4</td>
<td>.16</td>
<td>.21</td>
</tr>
<tr>
<td>Mean</td>
<td>.14</td>
<td>.25</td>
</tr>
</tbody>
</table>

Table 4: Overall and individual mean duration interval (in seconds) between 2-3 and 3-4 unstressed syllables in Spanish with an imitated English accent and in L2 English produced by Spanish speakers (S1-S4).

<table>
<thead>
<tr>
<th>Sp.</th>
<th>Spanish text in imitated English</th>
<th>English text in L2 English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-3</td>
<td>3-4</td>
</tr>
<tr>
<td>S1</td>
<td>.09</td>
<td>.21</td>
</tr>
<tr>
<td>S2</td>
<td>.10</td>
<td>.20</td>
</tr>
<tr>
<td>S3</td>
<td>.11</td>
<td>.19</td>
</tr>
<tr>
<td>S4</td>
<td>.14</td>
<td>.18</td>
</tr>
<tr>
<td>Mean</td>
<td>.11</td>
<td>.20</td>
</tr>
</tbody>
</table>

The results presented in Table 4 show that when Spanish speakers imitate an English accent, they produce the unstressed syllables between stresses quicker than when they speak L2 English. Actually, the mean duration intervals of an imitated English accent (0.11 sec. [between 2-3 unstressed syllables] and 0.20 sec. [between 3-4 unstressed syllables]) are closer to the mean duration intervals of L1 English (0.09 [2-3] and 0.18 [3-4]) than to those of L1 Spanish (0.14 [2-3] and 0.25 [3-4]) (see Table 3). Furthermore, the mean duration intervals of L2 English (0.16 sec. [2-3] and 0.24 sec. [3-4]) are more similar to those of L1 Spanish (0.14 [2-3] and 0.25 [3-4]) than to those of L1 English (0.09 [2-3] and 0.18 [3-4]). Given the reduced number of speakers, no statistical analysis could be carried out and more data...
needs to be gathered. However, as a first approximation, the results indicate that in imitated English-accented Spanish, Spanish speakers produce a rhythmic pattern closer to real English than when they speak L2 English. So, Spanish speakers are able to pick up the importance of English stresses when they imitate an English accent in Spanish but not in their own productions of L2 English.

3.2 The quality of unstressed vowels

This section examines whether the differences in the rhythmic patterns of the imitated English accent and those of L2 English pronunciation have an effect on the production of weak vowel schwa (/ə/) by Spanish speakers. Table 5 presents the percentages of weak vowel production in the two reading conditions according to the auditory judgements of the three phonetic specialists.

The results presented in Table 5 show that for all Spanish speakers the percentage of schwa production is higher in imitated English-accented Spanish than in L2 English. Two-tailed t-tests presented significant differences in all cases at p<0.01. These findings suggest that if speakers focus on the production of stressed syllables in English in connected speech rather than on the pronunciation of segments, the production of weak vowels may improve since they emerge more spontaneously.

<table>
<thead>
<tr>
<th>Sp.</th>
<th>Spanish text in imitated English</th>
<th>English text in L2 English</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>34%</td>
<td>12%</td>
</tr>
<tr>
<td>S2</td>
<td>41%</td>
<td>15%</td>
</tr>
<tr>
<td>S3</td>
<td>46%</td>
<td>20%</td>
</tr>
<tr>
<td>S4</td>
<td>32%</td>
<td>10%</td>
</tr>
</tbody>
</table>

4. DISCUSSION

This study has compared the production of rhythmic patterns and the weak vowel schwa by Spanish students of English in two scenarios: 1) in imitated English-accented Spanish and 2) in real L2 English. The results have shown that both the rhythmic pattern and the presence of the schwa are better in imitated English-accented Spanish than in the L2 speech. This indicates that whereas in an imitation scenario speakers are able to perceive and reproduce some of the more salient features of a language, in real L2 speech the production of L2 phonetic characteristics diminishes. These findings are relevant for phonetic teaching since they show that phonetic teachers can benefit from the students’ innate capacity for L2-accent imitation in their L1.

However, how imitation can be exploited for phonetics teaching and learning? Different scenarios can be considered depending on the immediacy of the imitation. Our first proposal is to start with the imitation of stored phonetics knowledge so as to see which phonetic features the students have innately picked up while listening to English and apply this knowledge to real L2 speech. Long stretches of speech (rather than words) with the same rhythmic structure in Spanish and in English would be presented to the students to read. It would be useful to include lexical items which are similar in the two languages, both in terms of segments and stress distribution, as illustrated in (2), so that students only pay attention to pronunciation matters. Students will be asked to produce the Spanish sentence first with an imitated English accent. The features observed in the imitation should be discussed in class with the teacher, paying special attention to stresses and weak vowels. Then, the students will be requested to adopt the imitated features to the English sentence and to their L2 subsequent productions.

(2) A*manda *coge-el *tren a *Brighton.
A*manda *takes the *train to *Brighton.

5. CONCLUSIONS

This study has examined the production of English rhythm and the weak vowel schwa by Spanish students of English in imitated English-accented Spanish and in L2 English. The results indicate that when Spanish speakers speak Spanish with an imitated English accent, they produce a rhythmic pattern closer to real English than when they speak L2 English. Similarly, the production of weak vowels is higher in imitation than in L2 speech. This corroborates the idea that rhythm is a key feature for a good imitation of English pronunciation with positive effects on vowel quality. The immediate implication of these findings is that using the rhythmic pattern of sentences as the starting point for L2 phonetics teaching can be especially useful for students of languages with different rhythmic patterns, such as, Spanish and English. Furthermore, rhythm is halfway between sounds and prosody. Thus, if rhythm becomes the core of phonetics teaching, the study of other areas of phonetics, such as intonation, will be benefitted since the intonation contours can only be explained in association with the metrical structure. Thus, this paper supports a top-down approach to phonetics teaching.
6. REFERENCES


---

1 This investigation has been carried out as part of the EMULANDO project financed by the Ministerio de Economia y Competitividad in Spain (Ref. no. FFI2014-59848-C2-1-P).
Intrusive tokens of aspiration in French learners’ L2 English

Christelle Exare

Laboratoire de Phonétique et Phonologie, CNRS (UMR 7018) / Paris3, Sorbonne-Nouvelle, France
christelle.exare@gmail.com

ABSTRACT

This paper addresses the production and the perception of ‘intrusive [h]’ in French learners’ L2 English (e.g.: I h ate pasta instead of I ate pasta). Based on the phonetic description of L1 and L2 English word-initial vowels, this paper provides three explanations and three teaching applications to deal with such phonologically illicit occurrences. First, the onsets of English words are phonetically analysed in three types of data: i) a text read by 8 native English speakers and 10 French learners of English, ii) spontaneous speech elicited from 25 French learners and iii) a perception test taken by 30 French-speaking students. The results suggest that a pause, some glottalisation or some aspiration are three processes that increase the time span between two vowels in a hiatus context. Second, three processes accounting for “intrusive h” are put forward. Third, some phonetic training is proposed.

Keywords: [h], intrusion, aspiration, English as a Foreign Language, French learners

1. INTRODUCTION

Intrusive tokens of aspiration are commonly referred to as “epenthetic /h/” [21]. Although rare [19], such occurrences are perceptually salient and often strongly stigmatised, particularly in proficient learners [8, 26]. Since such errors generate semantic inconsistencies (we h eat), they are supposed to deserve immediate and active correction. Some of the triggering parameters of intrusive aspirations have already been identified. Intrusive aspirations usually occur at stressed vocalic onsets. They are more common after a pause or a vowel, and when [h] can be found in the preceding context [21].

2. EXPERIMENTS

The onsets of some English words were analysed in three types of data: i) a text read by 8 native English speakers and 10 French learners of English, ii) spontaneous speech elicited from 25 French learners and iii) a perception test taken by 30 French-speaking students. The learners who took part in this study were all assessed at levels A2 or B1 of the Common European Framework of Reference for Languages. They all studied English in a classroom environment.

2.1 Production: Read speech

2.1.1 Protocol

The first experiment was a pilot study aiming at comparing hiatus resolution in L1 and L2 English. An ad hoc text including 20 noun phrases was read by 8 native English speakers and 10 fifteen-year-old female learners of English. The noun phrases consisted of the article THE followed by an adjective or a noun starting by a long vowel (/ɑː/, /ɔː/, /ɔː/, and /ɪː/). For example, one sentence was: “the autumn was so cold for the hawk and the eagle that they flew to the east.” The native speakers were from diverse origins. Three were from the United Kingdom. Three were from the United States. One was Canadian and one was from New Zealand. The productions were coded according to the type of vowel phonation chosen for hiatus resolution. Using Praat (Boersma 2001), four types of vowel phonation were acoustically coded: i) modal, ii) glottalised, iii) aspirated iv) other (e.g.: elision or unpronounced item). The rules established by [13, 25, 27] were followed to label the segments. In this paper, vowel-initial glottalisation acoustically corresponds to what is perceptually described as “hard attack” or “hard onset.” In a gestural approach [6], and as they characterise vowels here, the terms “aspiration” and “glottalisation” refer to two articulatory gestures that interfere with voicing. Glottalisation and aspiration can either precede the voicing gesture, or overlap it. The acoustic correlates of true English [h], which is often described as the voiceless counterpart of the following vowel, are numbered in figure 1 [25]. They are: a weakened first formant partly corresponding to the opening of the glottis (1), F2 and F3 clear and continuous transitions (2), a complete or incomplete interruption of voicing (3) and finally, aspiration noise in mid frequencies (4), and a drop in intensity (5).
In this French corpus of L2 English most tokens of intrusive aspiration could be segmented as [h]. In some erratic cases, however, /h/ was categorically perceived, although no real segment could be isolated. Instead, breathy vowels could be found. From an articulatory point of view, a breathy vowel has a double source: vocal fold vibration and aspiration noise. Hence the phrase “voiceless vowel” does not seem to be appropriate to describe such occurrences, as they are indeed produced with voicing, despite incomplete vocal folds adduction that results in aspiration noise.

### 2.1.2 Results

**First**, glottalisation is preferred to modal voice by both groups of speakers (natives and non-natives) at THE#V boundary, where # stands for a word boundary. However, the F0 curve tends to drop at the onset of the second vowel in native productions, while it goes up when an intrusive token of aspiration is produced in non-native productions. **Second**, three categories of learners emerge according to their choice of vowel phonation at V#V boundary. Most students produce both modal vowels and glottalised vowels in this context, but 5 of them significantly prefer glottalisation to modal voicing. 1 student produces 17 intrusive tokens of aspiration (in a total of 20 noun phrases) and no glottalisation at all. On the contrary, 2 other students never insert [h], while i) they utter glottalised vowels more than the others and ii) they always elide initial /h/ in contexts like “the harm,” “the hawk,” “the heart” and “the heater.” **Third**, glottalisation seems to be preferred by the native speakers who do not speak Southern Standard British English (see [11] for more detail on American glottal stops in V#V contexts). In this pilot study glottalised and modal voicing are two options equally chosen by British speakers. This is in line with [18]: vowel-initial glottalisation is not a forbidden option for native speakers of British English.

### Table 1. Type of phonation uttered by i) 3 native speakers of Southern Standard British English and ii) 3 native speakers of American English, at V#V boundary, in a THE#V context. For each speaker the same 20 noun phrases were analysed.

<table>
<thead>
<tr>
<th></th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>modal</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>glottalised</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>other</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

A chi-square test for independence compares the two distributions and categorical variables (in grey in the table above). The result indicates a significant association between the speaker’s linguistic origin and the type of phonation, as $\chi^2 (1, n = 117) = 6.69$ with $p=0.009$. So a relationship can be found between the type of phonation (“modal” vs “glottalised”) and the linguistic origin of the speaker.

### 2.2 Production: Spontaneous speech

#### 2.2.1 Protocol

25 French students (19 females and 6 males), aged 15 years old, were recorded during a semi-structured interview. The experimenter asked them questions about their studies, their hobbies and the trip they had taken two weeks before the recording took place. The goal was to elicit utterances like “I asked” or “she answered.” The acoustic data were collected through a headset microphone. On the whole, 160 minutes of L2 English casual speech uttered by 25 learners were recorded, orthographically transcribed, and acoustically analysed.

#### 2.2.2 Results

The main results concern i) intra- and inter-speaker variability, ii) the phonetic distribution of intrusive h, and iii) the role of the task. **First**, only 48 intrusive tokens of aspiration were identified and isolated at L2 English word onsets. They are produced by 12 (in 25) learners. Their frequency exhibits high inter- and intra-speaker variability. 2 students insert more than 10 intrusive aspirations in their recordings that approximate 10 minutes. 10 students do not utter intrusive aspirations at all during their interviews.

**Figure 1.** Spectrogram of *we hold*, pronounced by a female British speaker.
Intra-speaker variability can also be observed. 13 lexically similar utterances were carefully compared, like “we asked.” For each pair uttered by the same learner, one item had an intrusive glottal fricative at the onset of the vowel-initial word, while the second item was realised with some glottalisation. In other words, learners tend to choose either some glottalisation or some aspiration in order to pull away two vowels in a hiatus. A pause is a third option, that may, in addition, increase the distance between the two vowels. **Second**, in the present corpus, intrusive tokens of aspiration only surface i) in strict initial position or ii) after a vocalic sound, which may be an intrusive word-final schwa. When the word-final phoneme preceding the aspiration is a plosive (e.g.: *eat, trip, about*) or a fricative (e.g.: *is, practise*), a phonetic intrusive schwa emerges before illicit *[h]*. For instance, “*I love animals*” is pronounced [aɪləvə#haniməlz]. **Third**, the type of speaking task taken by the two different groups of participants seems to play a role in the emergence of phonetic tokens of aspiration. An independent-samples t-test indicates there is a significant difference (t = 2.13; p = 0.04) in the proportion of intrusions between read speech (M = 0.12, SD = 0.2) and casual speech (M = 0.01, SD = 0.02).

### 2.3 Perception

#### 2.3.1 Literature

The phoneme *[h]* is hardly perceptible in weak positions in Turkish and other languages [24]. Syllable-initial English *[h]* is visually well-perceived [20]. Although it may acoustically be confused with initial *[p]* by native and foreign speakers [9], English *[h]* is equally recognized by native speakers and French learners [22]. The learners’ phonetic representations of *[h]* seem to be at the origin of their particular treatment of English *[h]* [23].

#### 2.3.2 Protocol

A perception experiment was conducted in order to determine how French learners perceive: i) glottalised vowels and ii) word-initial English *[h]* when it is licit (produced by native speakers) or illicit (inserted by French learners). The hypotheses are i) learners perceive *[h]* adequately in both conditions (licit English *[h]* vs French intrusive aspiration), ii) initial glottalised vowels are perceived as *[h]*. 30 older and more advanced French students took the AXB test. 4 series of 20 stimuli were taken from i) the Aix-MARSEC Database [2], ii) the corpus of read speech described in section 2.1 and iii) the corpus of casual speech described in section 2.2. Each stimulus is a word followed by the first syllable of a second word starting either by a (glottalised vs modal) vowel or by (licit vs illicit) initial *[h]*. The syllabic boundary is that given by [29]. After a mock test the listeners had to determine whether they heard a vowel or *[h]* at the onset of the second word, in the 80 stimuli. 2400 responses were analysed.

#### 2.3.3 Results

The 30 listeners performed well. Initial vowels are perceived as vowels (82 % of correct answers) while *[h]* is generally identified as *[h]* (80 %). However, *[h]* is better identified when it is pronounced by English speakers than when it is pronounced by French learners: $\chi^2 (1, 1200) = 9.18$ with $p = 0.002$. Although a tendency to perceive glottalised vowels as *[h]* is noted, no significant effect of the type of phonation on the perception of the onset can be found: $\chi^2 (1, 1200) = 1.06$ with $p = 0.3$.

### Table 3. AXB perception test taken by 30 French listeners. The targets are L1 and L2 English vocalic and aspirated word onsets. Number and ratio of responses for each category. 4 categories of 20 stimuli are presented: i) modal vowels : (M) *[V]* ; glottalised vowels: (G) *[V]*; L2 English intrusive *[h]*: *[H]*; L1 English licit *[h]*: H.

<table>
<thead>
<tr>
<th>SOUND</th>
<th>Played</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M) V</td>
</tr>
<tr>
<td>V</td>
<td>512</td>
</tr>
<tr>
<td>H</td>
<td>88</td>
</tr>
<tr>
<td>%V</td>
<td>85.33%</td>
</tr>
<tr>
<td>%H</td>
<td>14.67%</td>
</tr>
</tbody>
</table>

### 3. DISCUSSION

#### 3.1 Proficiency level

No consistent relationship between the learners’ tendency to insert and drop *[h]* could be found. However, it is hypothesized here that a learner has a good command of the oral features if they:
i) preserve /h/ where the realisation of the phoneme is expected (the elision rate is below 10% of all possible cases).
ii) never insert intrusive aspirations (the insertion rate is below 4% of all possible cases).

Some future research may confirm whether the calculation of a H-ratio based on the rates of elision and insertion is a reliable level indicator.

3.2 Three explanatory processes

Illicit tokens of aspiration at word onsets in L2 English can be considered as occurrences of hypercorrection [14], which may result from three distinct, yet possibly complementary processes. First, some students may produce [ʔ] instead of [h] and /or [h] instead of [ʔ] because of an incomplete assimilation of the English [ʔ] ~ [h] contrast. Second, initial h-insertion (combined with final s-insertion) may signal some optional phonological repair, restoring French rhythm through typically CV#CV syllabic structures. Third, an intrusive aspiration may be an intrusive gesture of glottal opening [6], sometimes overlapping the voicing of the vowel (i.e. when [h] can hardly be segmented) or a glottal constriction gesture that fails to reach its target (i.e. an inchoative glottalisation). In fact, glottalisation and aspiration both correspond to glottal tension [28], although glottalisation correlates with a closed glottis and aspiration correlates with an open glottis. Confusion between the two modes of articulation may come from the learner’s lack of control over glottal gestures. From a phonetic point of view, word-initial intrusive tokens of aspiration in L2 English should be studied in relation to word-initial glottalised vowels.

4. TEACHING APPLICATIONS

Relying on multimodal input, the author’s teaching proposal aims at raising a learner’s awareness of i) glottal control for aspiration, glottalisation, and continuous modal voicing across word boundaries and ii) some syllabic specificities of French and English that make word boundaries potential stumbling blocks in French learners’ L2 English, iii) variability in the realisations of /h/. The framework for this set of activities is that of the weak-interface model [15, 16], which posits that instruction should combine explicit and implicit knowledge.

4.1 Multimodal input

Acquisition of languages is favoured by multimodal input. Relying on visual and auditory input can increase the linguistic and cognitive abilities of children with communication disorders [3]. Subtitling videos has been shown to improve the results of second-language learners in aural comprehension tasks [5, 10]. However, to train students in speaking, casual speech or repetition should be preferred to reading tasks, in order to minimise the risks of producing intrusive aspirations.

4.2 Training in glottal control

Students ought to be encouraged to distinguish between two processes to perform vowel-to-vowel linking: i) continuous voicing (corresponding to the uninterrupted vibration of vocal folds) as opposed to ii) glottalisation (corresponding to an abruptly closed glottis). Kinesthatic activities, like the placement of one’s hand near the glottis can be useful here. To make students perceive what glottalisation is, an analogy with French can be drawn. Glottal stops are usually produced at the beginning of exclamations like “Hein?” Also, the study of a popular song like “Hello” [1] can allow the teacher to introduce the contrast between continuous voicing, glottalisation and aspiration.

4.3 Syllabic specificities

French learners should be taught about syllabic structures. French prefers CV syllables while English favours CVC syllables [12]. French words ending with a phonological consonant tend to be uttered with a final --phonetic-- intrusive schwa in French [7] and in L2 English [17]. Hence students should be explicitly told that they do not need to add extra schwas to final consonants in L2 English.

4.4 Variation in the realisations of English /h/

French learners should also be taught about the prosody-driven variation in /h/ realisations. It is important to stress that English /h/ can be weakened in unstressed positions. In particular the opposition between HAVE (lexical verb) and HAVE (auxiliary) is a major difficulty for French learners. This gradient characteristic of /h/ could be represented as variable-sized symbols (e.g.: fences). Visualizing real-time spectrograms may also be an interesting tool to help students improve their pronunciation.

6. REFERENCES

Severe Communication Disorders in Children and Adults. Brookes.


---

1 For more detail on the acoustic cues for [h], the reader may want to refer to Kenneth Stevens (1998). Acoustic Phonetics. The MIT Press. Cambridge, Massachusetts. (Chapter 8.3.1.5)

2 The author wants to thank Timothy De Cotis, Lucas Drouhot, and two anonymous reviewers for their kind and helpful comments.
A GENRE APPROACH TO PROSODY: TEACHING INTONATION FROM A DISCOURSE PERSPECTIVE

Miriam Patricia Germani
Lucia Inés Rivas

Departamento de Lenguas Extranjeras, Universidad Nacional de La Pampa, Argentina
miriamgermani@humanas.unlpam.edu.ar
luciarivas@humanas.unlpam.edu.ar

ABSTRACT
In our teaching practice at university, we frequently notice the difficulties students face when accounting for the role of phonological features in oral texts. The materials available for teaching intonation seem to be insufficient to make students aware of the ‘semogenic value’ of phonological choices in the total meaning potential of the English language [6]. The lack of theoretical systematization showing the relationship between choices in the phonological, lexico-grammatical and semantic strata of language and their corresponding context may be responsible for our students’ helplessness in relating the meanings of particular choices to the global communicative value of the discourse in process. In search of a pedagogical tool to corroborate our intuitions, we approach the teaching of intonation through combining SFL descriptions of genre and the intonation theories available to us in the design of a series of classroom tasks.

Keywords: genre, suprasegmental and paralinguistic features, interpersonal and textual meanings.

1. INTRODUCTION
Teaching English intonation at university level in Argentina has traditionally been approached through specially designed course-books such as Brazil et al. [3], Brazil [4], Brown [5], O’Connor & Arnold [8], Tench [13], Wells [14]. However, the exercises and examples provided by these authors to demonstrate how the theory works are frequently limited to a series of sentences, isolated exchanges or brief fragments of texts with very little or no reference to the context of discourse. This leads students to analyse phonological features independently from other features of the language and from the global meaning of texts, which results in a rather poor analysis of how intonation plays a role in the building of discourse meaning. We understand that a theory which encompasses the choices at the phonological stratum together with those at other strata is necessary to be able to explain language in use in real situations. In search of such a theory, we participated in a research project at Universidad Nacional de La Pampa [9], exploring the realizations of language in choices at different strata, including the phonological one. We noticed that generic descriptions of language like the theory proposed by Martin & Rose [7] do not include this latter stratum. Thus, we embarked on the task of analysing samples of authentic language looking for correspondences between the lexico-grammatical and semantic choices described and their phonological realizations. Consistent correspondences have been found between the stratum of phonology and other strata in the language system in the expression of interpersonal and textual meanings [10].

In this paper we describe our teaching methodology in two phonetics and phonology courses at our university, the difficulties faced by students at the moment of accounting for phonological choices and their meanings in oral discourse and a pedagogical proposal to overcome these problems.

1.1 Courses characteristics and teaching methodology
English Phonetics and Phonology III and IV are the last two compulsory courses in the discipline at the teacher training college at Universidad Nacional de La Pampa. Placed in the third and fourth year of the course of studies, they deal with the teaching of suprasegmental and paralinguistic features, their perception and production in a variety of genres and text types. The theoretical basis used in these subjects are those habitually used to teach English intonation referred to above.

Our teaching practice is usually divided into two main stages. The first one deals with listening to and analysing oral texts, identifying phonological features and interpreting their meanings in context. The second one deals with the students’ production of oral texts in which they have to show their mastery of the phonological system to express meanings, emotions and attitudes. In this paper, we concentrate on the perceptive analysis corresponding to the first stage. In both courses, we expect our students to be able to explain how the speakers’ choices of suprasegmental and paralinguistic features construct meaning in the general linguistic context of the particular text. We guide this practice by showing that linguistic and
paralinguistic features such as tonality, tonicity, tone, pitch level, prominence, tempo, pause, etc. build interpersonal and textual meanings at the global level of the text in combination with other linguistic choices.

However, when students are asked to do this kind of analysis autonomously, they find difficulties in projecting meanings beyond the tone unit in which they occur. The reason for this may be the way the theoretical materials available to them show the meanings of intonation patterns, described above.

2. DIAGNOSTIC TEST

To determine up to what extent the students were able to identify phonological choices and explain their discursive meanings based on the knowledge previously acquired, we asked them to watch and listen to a fragment of oral language and to describe the intonation and paralinguistic features they could recognize, indicating the effect these had on the meaning projected by the speaker, in relation to the general communicative function of the text and to the speaker’s attitudes, intentions and emotions. We wanted to observe whether they could relate the phonological features to the organization of the message and the expression of interpersonal meanings.

2.1 Participants

The participants in this study were two groups of students taking the courses English Phonetics and Phonology III (22 students) and IV (9 students), who had already taken two or three one-academic-year courses in the area, respectively.

2.2 Procedure

We selected an excerpt of an interview to Steven Berkoff, a British actor, film director and playwright, in the BBC programme HardTalk [1], in which the speaker produced a story, in this case an Observation. We showed it to the students three times with the following instructions:

• Watch this extract from an interview with Steven Berkoff, a British actor, director and playwright, and observe what the communicative goal of this extract is.

• Now, pay attention to the phonological realization of his speech. What phonological features can you distinguish (high or low pitch, falling or rising intonation, paralinguistic features, etc) and how can you explain these choices in terms of Berkoff’s message (attitude, intentions, emotions).

2.3 Results

The results show that students could successfully identify the following suprasegmental and paralinguistic choices, ordered from the most frequently recognised to the least:

• tone choice (21 students), slow tempo (21 students), pitch level (20 students) tense or precise articulation (14 students), extension of segments in particular words (12 students), pause (11 students), prominence (9 students), pitch range and placing in the voice range (9 students), tonality and/or tone unit length (7 students), volume (5 students), voice quality -breathy or creaky voice- (5 students) and tonicity (1 student).

With respect to the meanings expressed by these choices, the students referred to the following:

• emphatic fragments or highlighted words, without specifying reasons for such emphasis (21 students), the speaker’s emotion (18 students), the speaker’s commitment to his message (7 students), the speaker’s presentation of information as new or shared (7 students), the speaker’s reflective state (5 students), the speaker’s intention to continue speaking (5 students), the speaker’s quotation of someone else’s words (5 students), the speaker’s intention to show contrast or to indicate the beginning or end of a paratone (4 students), the speaker’s social divergence or convergence (2 students).

These results show that most of the students could identify prosodic and paralinguistic features at a descriptive level. However, their explanations of the meanings at play in the interaction were very general or abstract or too restricted to a particular realization, without considering the wider discursive context or the communicative intention of this particular excerpt. The students applied theoretical concepts associating particular prosodic and paralinguistic meanings to particular instances without making reference to the context of situation of the real text or to the speaker’s aims. Figure 1 below is an example of the work done by one of the students.

As can be seen from the student’s comments, in 1, she recognised the phonological choice (high pitch level) in two isolated words and she associated the choice to explicit emotion, without making reference to what is going on in the text, which was in this case the speaker’s reaction to a particular memory. In 2, the student recognised prosodic and paralinguistic realizations in some words she identified as significant but she did not explain particular meanings in the context in which they appear. In 3, she recognised pitch movement and explained it with a label typically associated with this tone in
intonation manuals. Once more, she failed to recognise its meaning in this particular text.

Figure 1: Sample of a student’s analysis

Berlitz: Well, when I watched the BBC and I learned about drama, I'd never been to the theatre, and I was a little bit of a juvenile delinquent, and my mother said "come on, watch this, it's a fantastic play", I think they were doing a George Orwell ..., 1984 on television, live incredible, and I thought God this is fascinating, and in black and white and they had such amazing actors. And I was so ... drawn drawn1 to this, and that made me feel I wanted to attach myself to something in my life that would remind you and make it or... give an answer to why God put me on this earth. He put me on this earth to do something, to justify this extraordinary gift of life, and I don’t think you have to be... Shakespeare to justify your life, I mean you can just in fact marry and have beautiful children and give them life.
1The speaker uses high pitch to show emotional involvement and to illustrate the feeling he is speaking of.
2He stresses and extends very meaningful words.
3‘Drawn’ is said with a decisive fall which is at the same time emphasized by its repetition

3. DISCUSSION AND TEACHING IMPLICATIONS

Based on the results of the diagnostic test and on our observations in our daily teaching practice, we believe that it is possible to integrate phonological choices and communicative meanings following a ‘top-down’ approach. Genre theory as described by Martin & Rose [7] may offer a starting point for this kind of approach. As mentioned above, in our research project, correspondences between phonological choices and stages and phases in oral genres have been identified [10]. We think that is it possible to suggest a theoretical framework that combines this generic approach and phonological meanings in order to help our students in the production of speeches, storytelling and the use of intonation resources in oral discourse in general.

4. PEDAGOGICAL PROPOSAL

This proposal is oriented to students who can identify linguistic intonation choices based on the theories proposed by Halliday & Greaves, Tench and Wells [6, 13, 14] – tonality, tonicity and tone – and by Brazil [4] and Brazil et al. [3] – prominence, key and termination – and paralinguistic features as described by Brown [5].

Our proposal approaches texts from a global perspective focusing initially on the communicative function of the extracts and their consequent generic configuration. Following the teaching/learning cycle (TLC) proposed in Rose & Martin [11] for teaching discourse genres (Figure 2) we first set the context to guide students in a macro-analysis of the text through questions about the participants in the discourse, the identities and roles of speakers and listeners, the speakers’ communicative intentions, the setting and its characteristics, the text style – scripted or spontaneous, public or private – among others.

Figure 2: Teaching/learning cycle (Rothery 1994) in Rose & Martin [11]

4.1 First stage in the cycle: deconstruction

We start by deconstructing the text, identifying the different stages and phases and their lexico-grammatical, semantic and phonological realizations. For this stage we expose students to the audio version of Wilde’s well-known children story The Selfish Giant, taken from the web page ‘Storynory’ [12], where it is easy for them to recognize the typical stages in the narrative genre and their functions in the overall meaning of the text. As regards phonological choices we rely on Tench’s [13] functions of intonation to make them aware of the textual and interpersonal meanings construed by intonation. Focusing on textual meanings, we guide students to recognize the four stages of the genre, and how they are manifested in the language through lexico-grammatical choices – initial adverbials, change in subject participants, complex themes – and phonological realizations – high and low pitch, pause. In this story, 10 of the 12 stages start with high key, and they are separated by pauses of 3.5” in average, which triple those within stages.

As regards the interpersonal function of language, we analyse the language within stages to show how the attitudinal lexis construes evaluative meanings. These tend to be marked phonologically through features such as pitch, extension of segments, voice quality, among others. Once they have observed the phonological realizations of emotions in this story – high pitch for positive feelings and low pitch for negative ones, for example – we draw their attention to sections in the story where these phonological features evoke attitude in non-attitudinal lexis.
The following image shows our analysis of a fragment of the story corresponding to the beginning of a stage (Complication 2) and the role of phonology in the building of textual and interpersonal meanings. We use the software for speech analysis Praat [2] as a tool to visualize differences in pitch level (see image below).

**Figure 3**: Sample analysis of part of the story

Then the Spring came, and all over the country
high pitch at 435 Hz: beginning of stage ‘Complication 2’ and positive emotions
there were little blossoms and little birds. Only in the garden
blossoms/birds/garden: extension of segments- highlighting positive emotion
of the Selfish Giant it was still Winter. The birds did not care to sing in it
still winter: extended segments, low pitch (223 Hz), creeky voice setting.

_Birds:_ pitch reset (412 Hz)
as there were no children, and the trees forgot to blossom.

*extended segments.*

Once a beautiful flower put its head out from the grass,
_Flower:_ high pitch
but when it saw the notice-board it was so sorry for the children that it
so sorry: _extended_/low, high pitch
slipped back into the ground again, and went off to sleep. The only people
_only:_ pitch reset (384 Hz) - _constraint_
who were pleased were the Snow and the Frost.

_Snow:_ low pitch (214 Hz) _Frost_ very low pitch (196 Hz): _extended_/l/
Spring has forgotten this garden,’ they cried, ‘so we will live here
all the year round.’ _The Snow_ covered up the grass
with her great white cloak, and the Frost painted all the trees silver.
_Snow, Frost, all, trees silver:_ _extended._

Low pitch throughout and declination (253 to 195 Hz)

**Figure 4**: Praat image

4.3 Final stage in the cycle, independent construction

For this last stage, we assign the same task with a different story for students to work on their own. They do their analysis and present their results to the class to share their findings and receive feedback from classmates and teachers to later produce a final version of their work.

4.4 Preliminary results

For this presentation, we include an example of the results of one of the experiences with a group of students. We could observe that after a brief training session, students could start to perceive phonological choices in association with generic configurations and explain meanings grounded in the context. In this way, they could notice the use of high pitch and relate it to a structural function (beginning of stage) in combination with other lexico-grammatical choices (change of perspective/time/characters). As regards interpersonal meanings within a stage, they could associate the high pitch with the expression of positive emotions and attitudes in connection to both explicit attitudinal lexis and implicit evaluative meanings. They could also notice the combination of features producing a certain effect, such as low pitch, slow tempo and long pause indicating the end of a stage (textual meanings) and falling tone, low pitch and extension of segments within a stage realizing negative evaluation together with a particular wording (interpersonal meanings).

5. CONCLUSIONS

We have proposed a methodology to approach the teaching of phonological meanings from a top-down perspective. Even though this is work in progress and our conclusions are just preliminary, we are confident that a theoretical perspective which considers the text’s overall communicative function and establishes relationships among the choices at different strata of meaning can help our students to understand the semogenic power of prosody within the total meaning potential of language. A theoretical approach based on genre theory, such as the one proposed by Martin and Rose [7], which observes the purpose of the text as a semantic unit, can be a useful starting point.

6. REFERENCES

2013)


REFLECTING UPON STUDENTS’ PROBLEMS IN PHONEMIC DICTATIONS

Marina Grasso

Universidad Nacional de La Plata, Argentina.
gressomarina@yahoo.com.ar

ABSTRACT

Students of the Teacher education and Translation courses at the Universidad Nacional de La Plata, Argentina, receive a year training on the segmental features and another one on the suprasegmental features of English. Worried about the significant number of exams failed due to inaccurate transcriptions, we have qualitatively analysed the phonemic mistakes in students’ dictations with the aim of helping learners correct their transcriptions once the acoustic stimulus is no longer available. We believe that during the revision stage students can use resources like contextual clues, their command of the language and their knowledge about the phonemic system of English to correct their transcriptions. We provide instances of frequent problems made evident in their exams and attempt to account for them. We also suggest some possible ways of helping learners improve their transcription skills, giving ideas on how to approach the correction stage before handing in their written tests.

Keywords: phonemic transcriptions – dictation – revision stage

1. INTRODUCTION

Phonetics in the Teacher education and Translation courses at the Universidad Nacional de La Plata, Argentina, is covered in two academic years. The focus is first placed on segmental features (phonemic and allophonic) and on training learners to use the IPA for making transcriptions and taking down dictations. Students must pass an oral exam and a phonemic dictation in order to be promoted. During the second year, Brazil’s [1] discourse-pragmatic approach to the interactive significance of English intonation is introduced and features of intonation are marked in the written exam. It is the number of tests failed due to inaccurate phonemic transcriptions in this instance that motivates this presentation.

Iivonen [2] states that the learning of a sound requires the creation of a memory model, then used by the speaker to produce the sound- and, we add, to transcribe it as well. Among the possible sources of mistakes, we think that the students’ mental representation of the sound plays a crucial role and sometimes seems to weigh more than the acoustic information received while the dictation is delivered. By analysing students’ tests we would like to reflect on what type of mistake can be corrected or avoided once the incoming acoustic stimulus is no longer available. Whether or not learners have a distorted idea of a sound, they possess some tools that can be of use when checking their transcriptions. Some issues related to perception are taken into consideration but the focus is mainly on the correction stage when students revise what they have written before handing in their tests. We present ideas on what can be done to help students improve their transcription skills and on how to go through their transcriptions in an effective way while taking the exam.

2. TEST ADMINISTRATION AND STUDENTS’ RESULTS ANALYSED

Written tests during the second year consist of a phonemic and intonation dictation recorded and delivered at the language lab. First, a recording of the whole fragment made by the teacher of the course is reproduced for learners to get acquainted with the context of the dialogue. The next steps consist of playing the fragment separated into smaller chunks for students to take down the phonemic transcription. Time is then given to check through. Then, the recording is played once more to double check what has been done so far. Lastly, the same rendition is repeated five times so that the intonational features are marked according to the selections made by the teacher. Although practice of phonemic transcription has extensively been done in the first and second years, it is the aspect that brings about more problems to the students.

With the learners’ consent we have qualitatively analysed a total of 60 written exams
administered to three groups of 20 students each, who have finished their second year of Phonetics. As figure 1 shows, there is a small difference between the number of exams that have been passed (49.26%) and those that have been failed (50.74%). We can see that inaccurate phonemic transcriptions were the common denominator among most of the tests that were below the acceptable standard (89.23% divided into 82.6% who failed due to transcription only and 7.23% of students who did not reach the required standard due to problems with transcription and prominence. Prominence seems to be a difficult intonational feature to detect, since it is the cause of failure in the rest of the tests analysed: along with the 7.23% already mentioned, 7.27% evidence difficulties to mark prominence only and 2.9% to mark prominence and tones.

Figure 1: Percentages showing passes and failures

As said before, the primary focus of this presentation is on the revision stage of the written exam. With the auditory input no longer available, we consider that the main tools students can recourse to are the contextual information of the passage, their command of the language and their previous knowledge of the segmental features of English.

The students who take the exams have an English language proficiency level between B2 and C1 on The Common European Framework of Reference for Languages. They have already read the play from which the passage for the dictation is extracted. The play is also analysed in class and used as the basis for the oral test, which means that students have information about the context in which the selected fragment occurs. As regard information on segmental features, weak forms and rules for inflectional endings are particularly emphasised. Processes of assimilation are also covered but they are not tested in dictations.

We have grouped mistakes on the basis of the information and the knowledge students have to approach the correction stage.

3. STUDENTS’ MISTAKES

3.1 Neutralization, weak and strong forms

Although a great deal of attention is paid to weak and strong forms, it seems a challenging and difficult topic for our students, as many of the dictations reveal. As regards the perception of weak forms it might be important to mention that vowel reduction in Spanish involves slight centring movements [3] which do not cause Spanish vowels to lose their intrinsic quality [4]. This might hinder the perception of vowel reduction in English. However, we had assumed that their knowledge of the inventory of weak-forms would contribute to correcting inappropriate transcriptions. In all the examples inadequacies are presented in bold type and prominent syllables are underlined.

1. | ðeɪl̚ prɒbəblɪ hæv mæʃɪ.nz tə dəɪ tʃu.n | wɛn  
   | ʃeɪt kæn ɪmpɔːt ˈdem frəm əmərɪka |
2. | tjuː uː vəʊ məʊldɪst frendz hɪʊ tɪli ən klək |  
   | hæv ɔːlɪdɪ ɪznəm |

As cases 1 and 2 exemplify, the strong form of objective pronouns in unstressed positions and the weak form of ‘have’ when used as a main verb, or strong when used as an auxiliary are among the most common mistakes. These occurrences seem to suggest that students fail to take weak or strong forms as a feature of the utterance in context.

Segments 3 and 4 present words that do not have a weak form but that have been taken as if they did- a recurrent error, especially with the preposition ‘on’ and the adverb ‘then’.

3. | æɹ fɛlt lɛik tɪ | æɹt əːl | daʊnt ɡɑʊ ən ənd  
   | mædʒɪn |
4. | wæn deɪ | jʊl bi sɪktstɪ ʃæv | ənd ɵn ju ʃei ɲəʊ  
   | wɔt æt əm ɬɪˈnəm ɲəʊ |

In our opinion, this evidences an overgeneralization of the use of weak forms. Being aware of the existence of some prepositions that have a weakened and a strong form, students expand the use of weak forms to other prepositions that do not offer this possibility. As regards the adverb ‘then’, it might be associated with ‘that’ which, indeed, has both forms depending on whether it functions as a demonstrative or as a connector.

Feige & Bohn [5] hold that stress and vowel reduction are treated as independent phenomena by Spanish speakers. This may have a connection with the examples found in the dictations analysed. It appears to be a difficult task to internalize the relationship between weak and neutralized forms and the lack of prominence- as the use of ‘you’ and
‘he’ in their strong forms show in #5 and 6 respectively.

5. | ar kan tel bai ðə lok in jor at | ðet juyv hyd səmθəŋ |
6. | hiz ælwiz wɔntd bɔ:monli| to bi ðə laik pksfəd ðə pwsəbl |

Once more, this seems to suggest that students overlook the fact that neutralized and weak forms must be analysed in connected speech and that linking use and context can prevent mistakes.

3.2 Pronunciation rules

Strong emphasis is given to the rules for inflectional endings. However, many transcriptions indicate that students fail to apply them- as examples #7 and #8 prove in connection to ‘ed’ endings (‘seemed’ and ‘opposed’), and fragments #9 and #10 as regards ‘s’ inflections (‘feels’ and ‘cranks’).

7. | hi si:m kwat ð bi:muŋ bo: | 8. | ðet ri:praːtent tu: differnt pɔnts ðə vju: |
   ðætsi: apəʊz ðəl | 9. | ðæts wɔt ð fi: laik tm |
10. | ðæn ðən wɔn ðə ðəuəəz k İz kəɪm | ðə inspektə |

As regards the transcription of ‘seemed’ in case #7 we might say that the devoicing of the alveolar plosive in this context could have made the student take the lenis plosive for its fortis counterpart. But we cannot explain the other examples in the same way so we think that students easily forget to apply the rules if practice is interrupted.

We believe the two segments below reveal an overgeneralization of the –s ending applied in inappropriate contexts.

11. | ar dəʊnt keə təpənz wɔt hi iz | 12. | wi hæv ð niu vɔu z fənsəl həu |

There appears to be a tendency- especially with unfamiliar or infrequent terms- to treat words ending in the alveolar fricatives /s, z/ as if they were plural nouns, such as ‘tuppence’ in #11 and ‘Vice chancellor’ in #12.

3.3 Grammatical mismatches

Students’ misperceptions can result in grammatically incorrect versions. We think that their knowledge of the language could sometimes assist them in noticing these mismatches and providing the right word while carefully revising what they have written, whether or not the appropriate form has been perceived during the dictation of the sequences.

13. | æim səprɔuz ðə ju | 14. | dəktaʒ lɪdɪz bi:n ðə vərɪ səkʃəsfl dɪrektsər ðə edʒəkərʃən |

The command of the language required to make corrections might be different in cases #13, #14 and #3 as well (æi felt laik r| ðæt æəl]. We believe that the context in #14 provides more information that students can consider in order to provide the right transcription: not only does the verb (‘has been’) indicate the need of a subject in the singular form, but also the fact that one surname and one director of education are mentioned reveals that the speaker is talking about only one person. These extra clues should lead the student to remove the plural in ‘doctors’. Although these observations seem obvious, the examples analysed indicate that learners not necessarily apply their knowledge of the language while revising their transcriptions.

3.4 Contextual errors

There are always cases in which the wrong alternatives given by the students result in an incongruous segment.

15. | ðæf wɔn ðə aʊə medz | bɾɔ:gt in æə tɾəum ləkŋ laik ðæt | rimed wɔd həv ðə fi: | 16. | mɪsɪz lɪmdən | mæi hæzəndz sɪkstɪ fəv tədə [⋯] juyv həd səmθəŋ |
   ælfrɪd: | ðæts nɪnt fəz |
   mɪsɪz lɪmdən | wɔulz kəntn əfəd tə bi fəo |

In #15, it is evident that a maid would not bring ‘a train’ to the table and that what follows is unintelligible- the text being ‘Rene would have a fit’. The same happens in #16 where the text reads ‘Wives can’t afford to be fair’. Neisser’s ideas [6] of ‘analysis by synthesis’ extended by [7] help us understand occurrences like these. In Stansfield’s words [8] ‘during comprehension, the listener continuously synthesizes speech into “chunks” and formulates hypotheses about what is said in each. This process is known as analysis-by-synthesis. If the perceived speech matches the hypotheses, the meaning is understood. If not, the internalized expectancy grammar formulates a new hypothesis about the input heard’. ‘Train’ is a more frequent word than ‘tray’ which might even be unknown to some learners. As a consequence, the student’s grammar of expectancy [9] deviates him/her to formulate a new hypothesis from the original text,
which- having read the play- is, at least, an improbable option. It would have been interesting to ask the students who wrote these inconceivable options to help us shed some light on these cases, since we cannot venture a sensible explanation that accounts for them.

4. FINAL REMARKS AND SOME PEDAGOGIC IMPLICATIONS

The reasons that lead to making a particular mistake might not be as transparent as we would like them to be, but we claim that some inadequacies in phonemic transcriptions can be more easily corrected if students are helped to exploit the tools available to them to do so.

In order to improve their transcription skills, students should be encouraged to read passages in phonetic script and make transcriptions, since these practices can have a positive effect on their dictation abilities [10] and they can improve the learners’ pronunciation and phonetic understanding [2]. Error correction activities can help focus on specific aspects of segmental features that need to be reactivated.

Gómez Lacabex et al’s study [11] shows that specific training in English vowel reduction in lexical words has a positive effect on Spanish learners’ perceptual skills. Taking Lotto & Holt’s idea [12] that ‘learning is enhanced by hearing multiple speakers produce target phonemes in a training set’, we understand that implementing listening activities can contribute to present variance of these variables to develop students’ perceptual abilities.

We see the need to raise awareness of the fact that phonological variability is context dependent. More emphasis should be given to the message transmitted. Contrasting utterances with different highlighted ideas might help students link the use of weak or strong forms with different phonetic environments.

In connection to exams, and from the examples provided, it seems safe to conclude that learners approach the revision stage as a bottom-up process, by looking at individual sounds or words without treating the text as a whole. This can result in their unawareness of mistakes that appear to be so obvious once they get back their tests corrected by the teacher.

There are two moments during the administration of the test when time is allotted for revision. The first one is after the recorded text is played for students to take down the phonemic transcription of the passage. We believe at this stage the use of a top-down strategy can be of great help: learners can go through the text paying special attention to the meaning expressed in it. This may contribute to the identification of contextual errors (such as the inconceivable alternatives in #15 and #16) and of misuses of weak or strong forms (as in #1 and #2).

A second reading promoting a bottom-up strategy can then be suggested: students can go over the passage marking doubtful transcriptions to check more carefully when the recording is repeated (such as /təpənz/ in #11) and bearing in mind the rules for inflectional endings to avoid mistakes such as in cases #7 to #10. Grammatical mistakes, e.g. subject-verb agreement (such as #14) or even problems with tenses (as in | hav ə:ridi rizən| in #2) can be dealt with at this stage.

After the text is reproduced to mark intonational features, students are given time once more to check through. It can be advisable to ask learners to try to carefully read their dictations so that they can notice any other mistakes, e.g. /bɾaɡət/ in #15, where there seems to be a transfer originated in the spelling of the word that might be perceived if said out loud. With prominence marked, students can also revise the occurrence of neutralization and weak and strong forms at this stage, to correct mistakes such as #5 and #6.

This is a first approximation to some problems manifested in students’ dictations in phonetic script. Although they might not be significant in number compared to other inaccuracies found in exams, we believe they gain importance since they can be corrected or avoided- whether or not the appropriate form has been perceived during the dictation itself- if students are helped to exploit the knowledge and information they have during the revision stage. This analysis needs further development. There are other types of mistakes which also deserve attention and will be the topic of future studies.

4. REFERENCES


1 We have tried using authentic material and also exchanging recordings with other teachers who offer practical lessons to different groups but results have proved to be considerably better when learners were exposed to their own teacher’s recordings. Thus—although presented with non-native production— we considered it was fair to prioritize the students’ familiarity and attunement with their teacher’s speech.

ii All the examples are taken from dictations of fragments extracted from The Linden Tree, by J.B Priestley.
LSC INSTRUCTION ON L2 PRONUNCIATION DEVELOPMENT

S. M. Mohibul Hasan
BRAC University
mohibul@bracu.ac.bd

ABSTRACT
The current research paper examines the effectiveness of Learning Style Conscious (LSC) instruction through the utilization of a classroom experiment in an English as a Foreign Language (EFL) context. The study was a controlled treatment with both experimental (N=18) and control (N=10) groups. The learners in the experimental group received LSC instruction while the students in the other group were treated with placebo instruction only. Their pronunciation progress in the perceptual mode was measured by pretests and posttests, with the following conclusion. The experimental group benefited significantly from LSC instruction.

Keywords: LSC instruction, development, L2, perceptual pronunciation, classroom study

1. INTRODUCTION

Though Pronunciation Instruction (PI) has been an important facilitator of L2 acquisition, few studies experimented with the possible manipulation of learner cognitive style-relevant features in the actual treatments. One reason why primary studies often summarily ignored learner characteristics was probably because of the over-simplification of the causes of L2 learner-related errors by the proponents of Contrastive Analysis Hypothesis (CAH: [8]). According to the hypothesis, dissimilar L2 sounds are supposed to create relatively more difficulties to learners of an L1. Accordingly, language experts, biased to CAH, think that contrasting the phonological features of L1 X and L2 Y holds the holy grail of classroom pronunciation solution containing general predictions regarding the supposed difficulties of L1-X speaking learners of L2-Y (e.g. L1-British English speaking learners of L2-Greek). Nevertheless, multiple well-vetted theories (see Speech Learning Model [4]; Perceptual Assimilation Model [1]) in SLA already pointed to the inefficacy of language instruction based on CAH, which mistakenly considered similarities and differences as correlates of ease and difficulty [7].

In the context of Bengali EFL learners, publications spread out over decades (e.g. [6 and 11]) point to the sustained interest in classroom pronunciation development. However, the experimental nature of such studies has been a very recent trend in the region (for a typical study on the subject, see [10]). Additionally, most PI studies paid complete attention to only productive pronunciation development. There is yet to be a study specifically focusing on the perceptual pronunciation in the L2 classroom. The present study intends to fill this research vacuum, reporting on the LSC instruction of L2 English stop consonants in the perceptual mode with L1 Bengali speaking learners.

2. METHOD

2.1 Instructional target

Though theoretically English /b, d, g/ are voiced stops, in syllable-final positions the phones are almost-voiceless or completely voiceless [12]. For L2 learners, the only noticeable cue to the identification of syllable-final /p, t, k/ and /b, d, g/ is pre-fortis clipping and pre-lenis lengthening respectively. However, unless trained, L1 Bengali learners of L2 English do not usually hear for clipped and lengthened vowels preceding /p, t, k/ and /b, d, g/. As a result, learners’ perception of the six plosives syllable-finally has been reported a challenge. In the current study, identifying English /p, t, k/ and /b, d, g/ in syllable-final positions is considered one of the first baby steps towards the perceptual development of all stops in English.
2.2 Research question and hypothesis

After receiving 15 hours of LSC instruction (10 1.5-hour classes), will the experimental group register superior scores in their perceptual identification of the six English stop consonants in syllable-final positions?

The author of the study hypothesized that the intervention subjects would benefit from LSC instruction and would develop their perceptual pronunciation. If they were left untrained, their incapacity to recognize the L2 contrast could have caused them to continue having issues regarding their perceptual intelligibility [9].

2.3 Participants

The study employed a total of 28 participants (see Table 1). The students of the LSC group were final semester postgraduate students of a tertiary institution in Bangladesh. The control section consisted of third year undergraduate students from another institution. None of the participants reported any significant residence in any L1-English speaking country. Even though the curriculum levels of the two groups were different, the pretests pointed to their being equally matched on their actual levels of language exposure (see section 3).

Table 1: Participant Details

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Age</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>22.3</td>
<td>10</td>
</tr>
<tr>
<td>LSC</td>
<td>24.9</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>28</td>
</tr>
</tbody>
</table>

2.4 LSC and control instruction

According to Witkin [13], field dependent (FD) people take a holistic approach to processing language input rather than discreet analysis, which appeals to field independent (FI) people. The students in the LSC section were given a questionnaire regarding the classroom conditions and instructional features that they considered optimal for L2 acquisition. A careful analysis of the filled-in questionnaires convinced the instructor that most of the students were FD. Thus, the teacher designed LSC instruction specifically focusing on the need for integrating target forms within a meaningful context. More specifically, the LSC students were treated with pronunciation-focused tasks [3] integrated within debating activities. The activities required them to develop their assigned topics, which had the dual aim of challenging them intellectually and teaching them the target pronunciation forms. Here is a sample topic: “Noisy pups are not allowed in certain pubs” in Edinburgh.”

During the instructional sessions, the teacher intentionally hyper-articulated each target word (e.g. wet/wed with clipped/lengthened vowel respectively) at least two times. There was a pause of 4 to 6 seconds between two sequentially-produced target words. The students were not corrected during their presentations even if their pronunciations were not target-like. Rather, during the feedback sessions following presentations, the teacher specifically pointed out students’ non-target-like pronunciations (if any). The advantage of such activities over so called “drill and kill methods” [5] is that the latter type narrows down instructional focus on select pronunciation forms without any consideration of meaning in context, which--as per the questionnaires--most students of the LSC cohort would have found off-putting.

The pretests consisted of 50 tokens with syllable-final /p, t, k/ contrasting with /b, d, g/. The posttests consisted of items similar to (not the same) the ones the students attempted in the pretests. The tests consisted of items such as the following:

- Audio Stimulus: Bet (followed by)
- Answer Choices on Hard Copy Test:
  Choice 1: Bed; Choice 2: Bet; Choice 3: Beg

After listening to each audio stimulus, the students had to tick only one (the left, middle, or right token) of each triplet (three-alternative forced-choice).

3. RESULTS AND DISCUSSION
Overall, the participants in the experimental group benefited from the LSC treatment they received. In fact, they registered a mean difference of 18.06 (time 1 mean: 107.44; time 2 mean: 125.50). On the other hand, the control students suffered worsening in the posttests achieving a negative mean difference of 1 (pretest mean: 108.1 vs posttest mean 107.1; see Figure 1).

**Figure 1: Average Time 1 to Time 2 Scores of Control and LSC Groups**

Independent samples T-tests confirmed that there was no statistically significant differences \( (p = .82) \) between the experimental and control groups at the outset of the study. The \( t(26) \) and \( d \) values were also insignificant (-.23 and -.09 respectively). Therefore any positive gains by the LSC group in the posttests may be attributed to the treatment received.

Statistically significant differences were found in the posttests \( (p < .001) \). The \( t(26) \) was 6.35, with an effect of 2.5. Additionally, in order to assess whether the groups individually registered any gains, paired samples T-tests were also administered. The experimental group’s performance between time 1 and time 2 was also statistically significant, with \( t(17) = 38.04, p < .001 \). However, the control section suffered attrition registering a negative pre to post gain, with \( t(9) = -1.3, p = .22 \).

Because it was visibly clear that the LSC section outperformed the control section in the current study by more than 2 SDs--which indicates a large gain according to Cohen’s [2] benchmark--it may be concluded that the application of LSC instruction has substantial gains in the modification of perceptual pronunciation in a classroom context. The author singled out a couple of variables that were crucial in causing the instructional success of the LSC section. First, the treatment had its foundation on noticing and attention because learners tend to learn the most if their attention is directed to the target language. Second, as most students in the experimental section were field dependent, LSC instruction consciously catered to their learning preferences by embedding the target pronunciation forms in the debating activities. Even though the dominant practice in PI classrooms is the use of context-removed drills along with rule interpretation, the instructor withheld that kind of treatment for the current study. The practice of the L2 stops in a relatively real life context (e.g. debating activities) may have helped the learners produce the L2 stops in a target-like way. Additionally, LSC instruction may have also increased their subconscious grammatical knowledge regarding pre-fortis clipping and pre-lenis lengthening.

In future studies, pronunciation teachers may therefore try to generalize the effects of the current perceptual pronunciation experiment into the productive pronunciation mode, which is the more dominant sub-domain of PI.

**4. REFERENCES**


While the /b/ in the word “pubs” is voiced due to the following /z/, the same phone is almost voiceless in the singular word “pub” (with a syllable-final /b/). The students utilized both the singular and plural forms during their debating sessions. In the feedback sessions, however, the teacher consciously used the singular form only and hyper-articulated the clipped and lengthened vowels in “pup” and “pub” respectively.
Spectrogram reading as a tool to teach acoustic phonetics

Mirjam J.I. de Jonge

University of Amsterdam, the Netherlands  
m.j.i.dejonge@uva.nl

ABSTRACT

Any course in phonetics must cover acoustic phonetics. This paper presents an approach to the teaching of acoustic phonetics using spectrogram reading that has been used successfully for over a decade in an introductory university course. The method is easy to implement, allows students to learn quickly, and builds knowledge and skills that extend to other domains in phonetics and linguistics. Moreover, students tend to enjoy doing the exercises, which makes for an enjoyable classroom experience for everyone involved.

This paper outlines the educational context of our phonetics course and the position of the spectrogram reading module, how the materials are developed, how assessment is done, and which learning outcomes are achieved.

Keywords: spectrograms, acoustic phonetics, student engagement, material design, assessment

1. INTRODUCTION

Part of the introductory course in phonetics at the University of Amsterdam is learning to read spectrograms, which is used as a tool to teach acoustic phonetics. By studying spectrograms with the objective of ‘reading’ them, students are presented with visual cues to the acoustic properties of vowels and consonants in a structured way that allows them to learn many new things fast. Doing spectrogram reading means students apply their general knowledge of acoustics in a specific instance, and it allows students to get a sense of accomplishment very quickly, which is helpful in keeping students motivated and interested in the topic. Students generally find spectrogram reading exercises fun to do and feel they are learning a skill, rather than simply memorising properties.

Materials for spectrogram reading exercises are easy to create for an instructor, whether the material is meant to illustrate basic phenomena or to serve as test material. We feel many instructors of phonetics could benefit from integrating spectrogram reading exercises in their courses, and offer this paper as an example of how to develop a spectrogram reading module as part of a phonetics course.

2. EDUCATIONAL CONTEXT

The introductory course in phonetics for first-year linguists at the University of Amsterdam aims to provide students with basic knowledge of speech perception, production and acoustics, and knowledge of Dutch phonetics, and aims to teach basic skills in speech recording and acoustic analysis, phonetic transcription, and articulation. The initial part of the course covers four main topics: acoustics, transcription of Dutch, perception, and articulation. The second part covers a bit more typological ground, considering the sounds of the world’s languages and historical sound changes.

Dutch is the language of instruction and also the main language of study. Languages that are commonly taught in secondary education in the Netherlands (English, French and German) are also used often to illustrate phonetic phenomena. The instruction and assessment materials presented in this paper pertain mostly to Dutch phonetics, but the approach would be easily transferable to other languages.

The central concept of the whole course is the interaction between speakers and listeners, and acoustics is presented as the channel through which the interaction takes place. All material is taught using a learning-by-doing approach where students have to reproduce the phenomena discussed in their English textbook [2] for Dutch or another language they know well. In the first session students learn to record and inspect their own speech using the program Praat [1], which is used throughout the course.

Before the spectrogram reading module starts, the course covers measuring formant values from waveforms and spectrograms of vowels in single syllables, so students are familiar with the parameters of the spectrogram. The articulatory definitions of (pre)voicing and aspiration have also been introduced. The spectrogram reading module is coupled with the study of the Dutch sound inventory, including regional and social variation and correct use of IPA symbols, and the main goal of the module is to endow students with basic knowledge of acoustic concepts and the ability to apply their knowledge independently to perform phonetic analyses.
3. INSTRUCTION MATERIALS

3.1 General setting

Before every class, students complete assigned reading on the topic of the class, e.g. ‘plosives’ or ‘nasals’. Students are required to record their own speech and study the spectrograms to identify visual cues and compare these to the English examples in the textbook. They are also encouraged to seek out and study speech material of varieties of Dutch different from their own, using databases like the ‘Speaking Map’ of Dutch dialects [3], or YouTube [4].

Classes start with discussing homework, to explore which cues students discovered by themselves and which cues will require more elaboration from the instructor. When discussing homework and classroom exercises, spectrograms are both projected on screen and distributed on paper. The paper handout allows students to scribble notes and mark salient properties for their own reference, and the projection enables students to come up to the front of the class and literally point to the cues they identified. This tends to be more effective than letting students verbally note or describe what they are seeing in a spectrogram.

When the homework has been discussed, the relevant cues to the phenomenon under investigation are identified in spectrograms of a limited set of items with minimal variation that can be real words or pseudowords. A similar set can then be used for the first spectrogram reading exercise. Figure 1 shows an example of a demonstration or practice set for place of articulation cues.

**Figure 1**: A very limited spectrogram set to highlight cues to Place of Articulation. Syllables to be matched are [bap], [dat], [gak].

When students are just starting out with spectrogram reading, practice sets should always be labelled, i.e. there should be a list of words or syllables to be matched with the spectrograms. Such matching exercises allow students to use their knowledge of acoustic phonetics in both bottom-up and top-down directions. They can start with the spectrograms themselves to determine properties like number of syllables, presence of plosives, and presence of corner vowels, but they can also start with the labels and determine which properties they need to find at specific positions in the spectrograms to make a match. Unlabelled exercise material only allows for bottom-up analysis, which is great practice once the basics have been covered, but we suggest waiting a few sessions before introducing unlabelled practice sets, which are discussed further in section 3.6.

3.2 Vowels and variation

Vowels are presented initially in single-syllable words like [be:k] and [bit] and the spectrogram of the vowel is only inspected from 25% to 75% of the vowel duration, so transitions from or into the consonants are not given attention at this stage. Formants of monophthongal vowels are easy to identify visually, allowing students to learn which formant patterns are typically associated with the corners of the vowel space. We use the terms ‘bottom heavy’ ([uː]-like), ‘top heavy’ ([iː]-like) and ‘spread out’ spectrograms ([a]-like) for these patterns.

Diphthongal vowels provide a more diverse learning opportunity. Dutch has both true diphthongs and diphthongised tense vowels, and their realisations are a source of regional variation. Students have to determine whether their own high-mid tense vowels are diphthongised or not. Enforcing awareness of diphthongisation in vowels that are traditionally considered monophthongs in Dutch prepares for the transcription part later on in the course, and helps students improve their pronunciation of foreign languages; all students speak English, where diphthongisation is stronger than in Dutch, and many speak German and/or French, where diphthongisation is minimal.

A practice set for spectrogram reading that can be successfully completed based only on vowels requires careful selection of items. For English, the numbers one to ten would make a nice set.

3.3 Plosives: voice and place of articulation

Plosives are exceptionally easy to identify in a spectrogram, which is why they are the next sound class to be studied in our spectrogram reading module. First, students are instructed to look for ‘gaps’ in the spectrogram signalling the closure of the vocal tract, and they learn to identify prevoicing and aspiration. At this point in the course the different voicing contrasts in Dutch, German and English are also discussed.
Identifying place of articulation in plosives is less straightforward because of the different types of cues that are involved. Formant transitions are most clearly illustrated in conjunction with different vowels, so a practice set like in Figure 1 but with several different vowels should be used. Burst properties are best covered in conjunction with different fricatives (section 3.4).

3.4 Fricatives and recording settings

Recording and analysing fricatives for the first time is a good opportunity for students to explore the frequency range of the spectrogram. Standard settings in Praat [1] show the spectrogram only up to 5000 Hz, which is generally enough to identify vowels but cuts off a large part of most fricatives, particularly sibilants. Students have to identify the frequency range that is covered in the spectrograms of their own fricatives. At this point in the course, we also include an exercise to demonstrate the relationship between sample rate and the frequency range that is available for acoustic analysis. Students who have tried and failed to distinguish sibilants in 8 kHz recordings, both by ear and by spectrogram, are less likely to make the mistake of using low sample rates in field assignments in later stages of their studies.

Since fricatives and stop bursts share cues to place of articulation, spectrogram reading exercises should ideally include both. Once these properties are covered, students should be able to ‘read’ most matching exercises with ease.

3.5 Sonorants

Sonorants are the sound class with the trickiest cues to distinguish in a spectrogram, but their functional load in matching exercises can be fairly low once students are successfully recognising most vowels and obstruents. Place of articulation in sonorants can be compared to obstruents in carefully constructed sets. Because the cues to distinguish nasals and liquids from one another in a spectrogram are fairly subtle, it is advisable to make several practice sets with minimal pairs, ideally from different speakers.

In Dutch, liquids have a strong tendency to colour preceding vowels whereas nasals do so less, so students can learn to inspect the transition of a vowel into a sonorant to distinguish nasals from liquids. Nasals often undergo place assimilation, and studying the spectrograms helps make students aware of this automatic process, which will help them in the transcription part of the course.

3.6 Larger-scale matching and reading exercises

At the end of a class, students are given a spectrogram reading exercise that is a list of related words rather than carefully selected minimal pairs. Because students need to take into account several cues at once, this is more engaging than cue-specific exercises, and it offers a true ‘reading’ experience. Students often report a great sense of accomplishment from these exercises. The degree of difficulty of this exercise is determined first by whether labels are provided or not (see section 3.1), with labels allowing for more targeted inspection of the item. For unlabelled sets, the difficulty is mostly determined by the size of the semantic domain, e.g. ‘numbers under 10’ would be easier than ‘numbers’, and ‘weekdays’ would be easier than ‘European cities’. Figure 2 shows an example of a set of semantically related spectrograms.

Figure 2: A small set of related trisyllabic words from the semantic domain ‘sweet bakery items’. The easy version of the exercise would include the labels ‘speculaas’, ‘taai-taaiipop’, ‘boterkoek’.

Beyond operationalising the knowledge and skills the students are acquiring throughout the spectrogram reading module, larger-scale matching or reading exercises are useful to familiarise the student with regularities in the pronunciation and connected speech phenomena of the language under study, provided items are longer words or whole phrases. Vowel reduction and place assimilation have visible traces in a spectrogram, and a student who has had to recover this from a spectrogram is less likely to overlook these phenomena in transcription exercises at a later point in the course.

4. ASSESSMENT

Spectrogram reading exercises make up a substantial part of the written exam on acoustics, which is the first of four written exams in this course. Three types of questions are used, testing different aspects of students’ knowledge and understanding, particularly recognition and synthesis.

4.1 Assessing recognition

The simplest exam question to assess knowledge at the recognition level involves a single sound or sound class, where the student needs to identify a single property. Examples include: which of three
spectrograms shows a diphthongised vowel, or which of three spectrograms shows an approximant /r/. This exercise tests recognition of individual properties in a very constrained context.

A larger labelled matching exercise of 10-12 related words is used to assess to what extent a student is able to recognise multiple acoustic properties in a moderately constrained context, combining bottom-up and top-down analysis. This exercise usually includes items that are obviously different from other items as well as very similar pairs (e.g. [fia:lem] – [laamem], a pair that was often confused in an exam question involving Dutch cities). Students can rely on multiple cues to solve this exercise, which means getting a good score on this exercise does not require exhaustive knowledge, but a single mistake easily leads to multiple points lost because at least two items will be involved in a label matching switch.

4.2 Assessing synthesis

A very different type of exercise is to make students describe (and/or sketch) the expected spectrogram of a given syllable. This requires active knowledge of the acoustic properties of the sounds in the target syllable and uses only top-down reasoning. This type of exercise prepares students for thinking about de novo speech synthesis, which is addressed later on in the course, and allows differentiation between students who are explicitly aware of cues from students who just know enough to succeed at a matching exercise when there are bottom-up cues present. In effect, this exercise will show whether students have reached comprehension at the level of application or at the level of synthesis.

5. LEARNING OUTCOMES

The educational goals of our phonetics course as a whole were described in section 2. We use spectrogram reading primarily to create and reinforce knowledge and skills in the domain of acoustic phonetics, but as described in section 3, engaging with the material allows for plenty of opportunities to extend to other domains.

The basic body of knowledge and skills students acquire through spectrogram reading is called upon later in this course in more integrative assignments like designing a simple experiment. The awareness of spectral detail that develops through spectrogram reading allows students to formulate specific experimental manipulations. Particularly further study of language variation and perception invite the operationalisation of acoustic phonetic knowledge and skills.

The learning outcomes of the spectrogram reading module also benefit students beyond their further studies in linguistics. Because students gain hands-on experience with Praat software, the learning curve in subsequent courses where the program is used is leveled, and because students learn to record and analyse their own speech and study specific phonetic details, they can work independently to improve their pronunciation in foreign languages as they have also been made aware which properties are particular to Dutch.

In sum, using spectrogram reading as a tool to teach acoustic phonetics is an engaging and effective way to cover a lot of ground on basic acoustic concepts while also building practical research skills. It is easy to develop instruction and assessment materials for, and it provides a range of learning outcomes that can be built upon in further studies of phonetics, linguistics and language learning. We highly recommend making it a part of any introductory phonetics course.

6. ACKNOWLEDGEMENTS

The materials for this course have been continuously developed over the past decade by several instructors of the course: Titia Benders, Paul Boersma, Silke Hamann, and Karin Wanrooij, in addition to the author.

7. REFERENCES

LEARNER PERSPECTIVES ON PRONUNCIATION FEEDBACK

Pekka Lintunen, Aleksi Mäkilähde and Pauliina Peltonen

University of Turku, Finland
pekka.lintunen@utu.fi; aleksi.makilahde@utu.fi; pauliina.peltonen@utu.fi

ABSTRACT
Corrective feedback is an essential part of L2 pronunciation instruction. The type and source of feedback may vary depending on the context. Most earlier studies suggest that corrective feedback facilitates learning in general, but there are few studies that focus especially on the effectiveness of corrective pronunciation feedback.

This study focused on advanced learners of English and on their opinions and preferences regarding pronunciation feedback. We asked a group of participants (N=46) to respond to Likert-scale statements on pronunciation feedback. Our results suggest that pronunciation feedback is found very useful and teacher feedback is preferred. Opinions towards peer feedback are more varied but mostly positive. The variation in general attitudes and emotions towards pronunciation feedback is possibly due to individual learner characteristics. The results also suggest that providing feedback is a skill that should be practised during teacher education to increase prospective teachers’ confidence and competence in providing feedback.

Keywords: corrective feedback, advanced learners, pronunciation teaching strategies

1. INTRODUCTION
There is overwhelming evidence that pronunciation instruction facilitates the learning of second language (L2) pronunciation skills [9]. An essential part of such instruction is corrective feedback (CF). CF refers to explicit or implicit acknowledgement that the learner’s L2 performance has contained an error [e.g., 6, 13], and it may be formulated with the aid of metalinguistic information and terms. Through CF, learners are believed to notice gaps in their knowledge and incorrect patterns in their behaviour [11]. Combined with practice, CF can help learners increase the accuracy of their performance. Furthermore, when learners’ self-evaluation skills improve, their chances for autonomous learning increase.

CF can vary depending on who provides it. In formal learning contexts, learners’ errors can be corrected by teachers, peers or the learners themselves [14]. There are many contextual factors determining which strategies benefit the learner the most. As teachers are considered competent experts and authorities in their fields, feedback from teachers is valued and found useful. It should be borne in mind that pronunciation is often considered to be very intimately connected to one’s identity, and pronunciation teaching and feedback may also cause negative feelings for the learners [1, 5, 12]. However, learners might find feedback from peers less threatening and experience less anxiety in learner-to-learner discussions. To provide good peer feedback requires that the pronunciation and analytical skills of the learners are satisfactory for the target level in question.

Moreover, analysing learners’ pronunciation skills and providing CF based on the analysis is very time-consuming for teachers [14]. If learners are advanced and well aware of the phonetic features of the target language, they should be able to provide CF and receive it from their peers fairly accurately and efficiently, while the teacher can monitor the discussions and focus on the most common errors or errors that may have been ignored in peer feedback. Considering large group sizes and limited contact-teaching time, this makes pronunciation teaching more efficient, while ensuring that learners still receive CF on their performance.

The types and effectiveness of feedback have been studied, but earlier studies have seldom focused on pronunciation feedback in particular [6, 8, 9]. Earlier studies have also suggested that pronunciation errors might be more difficult to correct on the basis of feedback than errors in grammar [2]. Teachers’ and learners’ perceptions of the most useful CF may also differ [4]. Finally, providing (and receiving) feedback is a skill that can be practised: after CF training, university-level learners have been shown to be more confident and willing to provide peer feedback [10].

In L2 teacher education it is essential that pre-service teachers have an opportunity to consider how to give feedback that best supports learning. In foreign language contexts, pronunciation teaching for pre-service teachers is often geared towards their own pronunciation skills, whereas aspects related to pronunciation teaching can more often be overlooked [3]. In consequence, teachers gain experience as learners but their didactic abilities might not improve. Novice and experienced teachers have been found to differ in their perspectives on the importance and effectiveness of different types of feedback: novice
teachers are influenced by their own learning experiences, whereas experienced teachers are able to analyse the role of CF in the particular learning context more thoroughly [7].

The purpose of this study was to investigate the learner perspective on corrective pronunciation feedback. We focused on the perceived usefulness of feedback (section 3.1), affective attitudes towards feedback (section 3.2) and the preferred sources of feedback (section 3.3).

2. METHODOLOGY

For the purposes of this study, we focused on a group of university learners of English (N=46) who responded to statements on pronunciation feedback with a 5-step Likert-scale (strongly disagree – strongly agree). In the results, we have combined steps 1 and 2 as well as 4 and 5 to give the proportion of participants disagreeing and agreeing, respectively. The participants were first year students of English, most of them (84%) women. The median age of the participants was 20 (range 19–46). The participants were about to finish an obligatory course on English pronunciation. English was a major or minor subject in their MA degree. All participants did not aim at a career in teaching, but it was an alternative they could choose later during their studies.

During the BA-level course the participants had frequently monitored their own production, received individual and group feedback from a teacher, and practised giving feedback to other students. The participants attended different teaching groups, but all groups were taught by non-native speakers of English. The pronunciation course included segmental and suprasegmental pronunciation exercises, ear training and phonetic transcription exercises. In this study we focused on learner attitudes towards pronunciation feedback.

3. RESULTS

3.1 Usefulness of pronunciation feedback

In general, pronunciation feedback was considered very useful. The participants agreed almost unanimously over the facilitative awareness-raising effect of feedback. Corrective pronunciation feedback supports learning by making learners more conscious of their pronunciation and drawing their attention to aspects they may have overlooked before. The participants also objected to the negatively phrased statement on the uselessness of feedback almost unanimously. In addition, more than half of the participants reported that they could not learn how to pronounce English without feedback from a teacher (see also 3.3). The statements related to the usefulness of pronunciation feedback are listed in Table 1.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Ave.</th>
<th>Agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving my pronunciation is easy when I get feedback</td>
<td>3.5</td>
<td>57%</td>
<td>15%</td>
</tr>
<tr>
<td>Giving feedback to a fellow student helps my own pronunciation as well</td>
<td>2.8</td>
<td>26%</td>
<td>37%</td>
</tr>
<tr>
<td>I could learn how to pronounce English without feedback from my teachers</td>
<td>2.4</td>
<td>11%</td>
<td>52%</td>
</tr>
<tr>
<td>I don’t find pronunciation feedback useful</td>
<td>1.1</td>
<td>0%</td>
<td>98%</td>
</tr>
</tbody>
</table>

Most participants thought that improving one’s pronunciation is easy after receiving feedback, but the number of unsure answers was also high with this statement. This could reflect the fact that, in general, improving one’s pronunciation skills was not considered easy. Finally, we also asked the participants to consider whether providing peer feedback is useful for their own pronunciation; the responses were fairly evenly divided between learners agreeing, disagreeing or not being sure.

3.2 Affective attitudes towards feedback

Most participants were happy to receive pronunciation feedback. It should be pointed out, however, that there were a number of learners who did not like receiving pronunciation feedback. This might reflect individual differences in self-confidence or anxiety levels [see 1, 5]. Preferences for certain types of CF or learning situations may also affect learner opinions and explain variation. A similar effect stemming from different learner characteristics can explain the fairly evenly distributed responses to statements regarding the effect of feedback on self-confidence and feelings of anxiety or shame. However, negative feelings were not usually increased by the worry that fellow
Most participants have been participants mostly pronunciation skills. The responses were fairly evenly majority did not like to comment on their friends’ considered focused on peer feedback used to receiving and giving feedback on grammar. An emphasis placed on grammar during earlier causing to be more ashamed than grammar feedback. In addition, only 22% of the participants were of the related to pronunciation feedback are in Table 2.

Table 2: Affective attitudes

<table>
<thead>
<tr>
<th>Statement</th>
<th>Ave.</th>
<th>Agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I like to receive pronunciation feedback</td>
<td>4.0</td>
<td>65%</td>
<td>13%</td>
</tr>
<tr>
<td>Giving pronunciation feedback is challenging</td>
<td>3.8</td>
<td>67%</td>
<td>7%</td>
</tr>
<tr>
<td>Receiving feedback increases my self-confidence</td>
<td>3.1</td>
<td>30%</td>
<td>22%</td>
</tr>
<tr>
<td>I feel anxious when I get pronunciation feedback</td>
<td>3.0</td>
<td>43%</td>
<td>39%</td>
</tr>
<tr>
<td>I feel ashamed of my mistakes when I receive pronunciation feedback</td>
<td>2.9</td>
<td>35%</td>
<td>41%</td>
</tr>
<tr>
<td>I get nervous when I have to comment on someone’s English.</td>
<td>2.9</td>
<td>35%</td>
<td>39%</td>
</tr>
<tr>
<td>I feel confident when giving pronunciation feedback</td>
<td>2.5</td>
<td>13%</td>
<td>46%</td>
</tr>
<tr>
<td>I feel more ashamed when I am given grammar feedback than pronunciation feedback</td>
<td>2.4</td>
<td>22%</td>
<td>54%</td>
</tr>
<tr>
<td>Giving pronunciation feedback is easier than giving feedback on grammar</td>
<td>2.4</td>
<td>24%</td>
<td>59%</td>
</tr>
<tr>
<td>I like to comment on my friends’ pronunciation skills</td>
<td>2.1</td>
<td>13%</td>
<td>74%</td>
</tr>
<tr>
<td>When I receive feedback I worry that the other students might hear it</td>
<td>2.1</td>
<td>15%</td>
<td>76%</td>
</tr>
</tbody>
</table>

In addition, only 22% of the participants were of the opinion that pronunciation feedback makes them more ashamed than grammar feedback. This means that for most learners pronunciation does not appear to be more intimately connected to their identities and causing stronger emotional reactions. Most participants did not find pronunciation feedback easier to give than grammar feedback. The opinion that grammar feedback is easier to give can stem from an emphasis placed on grammar during earlier education. Accordingly, learners might also be more used to receiving and giving feedback on grammar.

The questionnaire also contained statements that focused on peer feedback. Most participants considered giving feedback challenging, and a clear majority did not like to comment on their friends’ pronunciation skills. The responses were fairly evenly distributed when it came to nervousness when giving feedback, and almost half of the participants did not feel confident when giving feedback. This lack of confidence can mean that the participants had not practised giving feedback before university, and would also require advice on how to give the kind of corrective pronunciation feedback that is appreciated by other learners and best supports learning.

3.3 Sources of feedback

Our questionnaire also contained more specific statements on the preferred sources of pronunciation feedback. On the whole, the participants mostly disagreed with these statements, which also reflects the wording. Based on the responses, the participants trust pronunciation feedback from both teachers and native speakers. The assumption is that native-speaker teachers during earlier education have been rare, and at the time of the questionnaire the participants were taught by a non-native teacher. Most participants reacted neutrally to the statement that native speaker feedback would be preferable. These statements are listed in Table 3.

Table 3: Sources of feedback.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Ave.</th>
<th>Agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>It’s difficult to accept pronunciation feedback from a person whose pronunciation is weaker than mine</td>
<td>3.4</td>
<td>54%</td>
<td>26%</td>
</tr>
<tr>
<td>I’m more relaxed when I receive feedback from a teacher than from fellow students</td>
<td>3.1</td>
<td>33%</td>
<td>28%</td>
</tr>
<tr>
<td>I would rather receive pronunciation feedback from native speakers</td>
<td>2.8</td>
<td>18%</td>
<td>30%</td>
</tr>
<tr>
<td>When I listen to my own pronunciation, that’s the best feedback I can get</td>
<td>2.4</td>
<td>9%</td>
<td>59%</td>
</tr>
<tr>
<td>I do not trust feedback from a fellow learner</td>
<td>2.3</td>
<td>15%</td>
<td>61%</td>
</tr>
<tr>
<td>It’s better to receive feedback from fellow learners than teachers</td>
<td>2.1</td>
<td>4%</td>
<td>70%</td>
</tr>
<tr>
<td>I do not trust feedback from a native speaker</td>
<td>1.6</td>
<td>0%</td>
<td>91%</td>
</tr>
<tr>
<td>I do not trust feedback from a teacher</td>
<td>1.1</td>
<td>0%</td>
<td>98%</td>
</tr>
</tbody>
</table>

In contrast, a clear majority was of the opinion that they prefer feedback from teachers to feedback from fellow learners. 9% thought that listening to their own pronunciation gives them the best kind of feedback.
on their skills. Most learners found feedback from weaker learners difficult to accept, but 26% of the participants disagreed with this statement, implying that weaker learners can also give feedback to their more proficient peers. When it comes to feelings of anxiety, the participants were fairly evenly distributed as to whether feedback from a teacher or a fellow student is less stressful to receive: 33% considered feedback from a teacher less stressful and 28% were of the opposite opinion. The rest reacted neutrally, which means that they did not think that the source of feedback had any effect on this matter. This is another indication of the fact that learner preferences vary when it comes to corrective pronunciation feedback.

4. DISCUSSION AND CONCLUSIONS

In our study, learner opinions on corrective pronunciation feedback were found very positive. The statements related to the usefulness of pronunciation feedback were often unanimously agreed with: learners considered corrective pronunciation feedback useful and thought that it facilitates learning. This suggests that CF has a central role in supporting the learning process in formal learning contexts. However, only 26% of the participants thought that giving peer feedback helped their own pronunciation.

Affective attitudes towards pronunciation feedback varied, reflecting the role of differing learner profiles. Despite the generally positive attitudes, our results suggest that a large proportion of our participants felt anxious or ashamed when receiving CF. Moreover, most participants did not feel confident when giving pronunciation feedback. These results lend support to earlier findings on the negative experiences caused by pronunciation teaching [1, 5] and show that university-level learners are not ready to give pronunciation feedback without practice [7, 10].

When the sources of feedback were considered, the teacher was regarded as the best source of corrective pronunciation feedback. The native language of the teacher was not, however, viewed as a decisive factor. Considering that our participants were advanced university-level learners of English, the emphasis placed on teacher feedback may be even somewhat surprising, but the explanation might lie in the fact that pronunciation feedback has been uncommon or unsystematic prior to this level. University-level learners are assumed to be quite autonomous, but our results suggest that teacher input is clearly needed in pronunciation feedback.

On the other hand, even though teacher feedback was preferred, our results suggest that many learners find receiving peer feedback less stressful. This implies that a combination of both would be ideal. Furthermore, individual learner characteristics may determine which source of CF is most successful. The proficiency levels of learners need to be fairly even as the majority of participants thought that CF from less proficient learners is difficult to trust.

The results related to student confidence are especially thought-provoking. Lack of confidence may be related to the learners’ own pronunciation skills: less proficient learners may find providing CF challenging, especially if they have to provide feedback to a more proficient learner. Although our participants had practised giving feedback during the course, they may feel that practice has been insufficient.

It is important to address the possible negative feelings that pronunciation teaching and feedback may cause. Awareness-raising can help pre-service teachers to become more confident and competent when providing feedback. Furthermore, it can also prepare them for dealing with their own students’ anxieties in the future. Anxieties and insecurities can be lowered through practice, and therefore it is also important to practise giving and receiving feedback in different peer group combinations to get used to different kinds of learners during teacher training [cf. 7].

Although based on a limited sample, our study has illustrated learners’ views on pronunciation feedback from different perspectives. While many of the participants of the study will become teachers, it should be noted that not all of them will choose a career in teaching. Nevertheless, they do all aim for a career as experts on the English language. One part of this expertise is to be able to comment on other speakers’ English pronunciation skills. Often university-level L2 education aims at improving learners’ L2 proficiency without any explicit emphasis on simultaneously learning didactic aspects [4]. Our results lend support to suggestions that in L2 teacher education pronunciation teaching methods should be an essential part of the curriculum.

The next objective of our project is to look into the practices of pronunciation feedback and to investigate how peer feedback is given and which aspects are given the most attention (e.g., the proportion of segmental and suprasegmental issues). It is important to compare learners’ beliefs with their practices. As earlier research has shown that providing feedback is a skill which can and needs to be practised, it is essential that pre-service teachers are given these opportunities during their studies.

5. REFERENCES

results of a preliminary study. Research in Language 11. 57–79.


IDENTIFICATION OF MANDARIN HIGH-LEVEL TONE AND HIGH-FALLING TONE BY VIETNAMESE LEARNERS

Liu, Hui-chuan
Kainan University, Taiwan

ABSTRACT
This paper explores the identification of Mandarin Chinese high-level tone and high-falling tone by Vietnamese learners of Mandarin Chinese. As there is no high-falling tone in Vietnamese, Vietnamese learners encounter great difficulty in learning this Mandarin tone, which further influences the acquisition of the high-level tone. In this study, a listening test is conducted to test advanced Vietnamese learners’ distinction among Mandarin tones, specifically the high-level tone and the high-falling tone, and to explore the reason accounting for the confusion between these two target tones. The results show that advanced learners still have problem in distinguishing Mandarin high-level tone and high-falling tone. Different from Mandarin native speakers who distinguish these two tones by pitch slope, Vietnamese learners seem to make the distinction based on syllable duration. The adoption of wrong phonetic cue for judgement accounts for their confusion between the two tones.

Keywords: Mandarin, tone, second language, Vietnamese learners

1. INTRODUCTION
Corder [1] pointed out that the systematic error patterns learners make in learning a second language provide a window to explore which learning stage learners are at and what strategies learners employ to acquire the target form. It is reported that Mandarin Chinese (MC) learners encounter more difficulty in learning tones than segments [2,3]. Research on Vietnamese-speaking learners of MC consistently reveals a significant mutual-substitution between MC high-level (H-level) tone and high-falling (H-falling) tone [4,5,6]. Research on the L2 learners’ performance of MC tones mainly focused on beginning learners [7,8,9]. Error patterns uncovered from such data could be a temporary phenomenon existing at a certain learning stage and distinguishing later. The performance of learners with language fluency and long learning experience is comparatively stable. Hence, this study investigates the performance of advanced VT learners on Mandarin tones for two purposes: first, to examine if the substitution between MC H-level tone and H-falling tone still exists in advanced learners’ performance, and second, to explore what learning strategy or what phonetic cues learners adopt to make the distinction between these two tones.

2. LITERATURE REVIEW
2.1 The tonal systems of MC and VT
Both MC and VT are tonal languages, in which every syllable carries a tone, and words with identical segments but different tones differ in meaning. There are four lexical tones in MC. The tone name and the pitch shape are listed as follows:

Yinping: high-level tone, marked as [ ̈ ]
Yangping: high-rising tone, marked as [’]
Shangsheng: low-dipping tone, marked as [’]
Qusheng: high-falling tone, marked as [’]

There are six tones in VT. The tone name and the pitch shape are listed as below:

Ngang: mid-high level tone, represented with no mark
Sắc: high-rising tone, marked as [’]
Huyên: low-falling tone, marked as [’]
Hỏi: low-rising tone, marked as [’].
Ngã: high-broken-rising tone, marked as [’].
Nằng: low-broken tone, marked as [’].

The level tone ‘ngang’ in VT is similar to the MC H-level tone but slightly lower in pitch. The high-rising tone ‘sắc’ is similar to the MC high-rising tone. The low-rising tone ‘hỏi’ is similar to one variant of the MC low-dipping tone. The only MC tone that does not have similar correspondent in VT is the H-falling tone. Although there is also a falling tone ‘huyên’ in VT, this tone starts at a low pitch and the pitch descends slowly with a mild slope. Differently, the MC H-falling tone starts at a very high pitch and descends with a large slope. Therefore, the H-falling tone is a novel tone to VT learners.
2.2 Literature on VT learners’ performance on MC tones

Previous research on the performance of MC tones by VT learners consistently reveals learners’ difficulty in learning the H-falling tone and a mutual substitution between the H-level tone and the H-falling tone [4,5,6]. All the studies attribute this learning problem to the influence of L1 as there is no tone with a large falling slope in VT, but inquiries such as how the L1 system affects VT learners in identifying the H-falling tone and why learners reveal confusion between the H-level tone and the H-falling tone have not been studied yet. Without understanding the cause of the errors, it is hard for teachers to help learners acquire these tones effectively in class.

3. METHODOLOGY

3.1 Experiment design

A listening test was conducted to collect data for analysis. The test was composed of mono-syllabic and disyllabic words with four tones evenly distributed. Each tone appeared 21 times in total. The stimuli were recorded by a native Taiwan MC speaker.

3.2 Subjects and data collection

Seven VT college students currently studying in Taiwan were recruited as paid subjects. All students majored in MC and have studied MC for at least 3 years. In the test, the stimuli were played by computer automatically. All items were repeated twice following 6- to 8-second silence for subjects to write down the words they heard with Hanyupinyin and the tone mark. Hanyupinyin is a spelling system of MC composed of 26 alphabet letters.

4. RESULTS

4.1 Analysis of the accuracy rate

The results of the test in terms of the accuracy rate are shown in Table 1. Given the accuracy rate of 90% as a threshold of mastery, the accuracy rates of the rising tone (99%) and the dipping tone (97%) indicate that all subjects have no problem identifying these two tones. This result is expected and attributed to the L1 influence as there are also similar tones in VT.

<table>
<thead>
<tr>
<th>Tone</th>
<th>Level tone</th>
<th>Rising tone</th>
<th>Dipping tone</th>
<th>Falling tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokens</td>
<td>125</td>
<td>145</td>
<td>143</td>
<td>111</td>
</tr>
<tr>
<td>Acc. rate</td>
<td>88%</td>
<td>99%</td>
<td>97%</td>
<td>76%</td>
</tr>
</tbody>
</table>

The accuracy rate of the H-level tone is slightly lower than 90%. There are 2 subjects not reaching the mastery requirement. The accuracy rate of the H-falling tone is less than 80% and more than half of the subjects do not reach the mastery requirement. Errors of the H-level tone and the H-falling tone display mutual substitution of these two tones. This also conforms to previous studies.

The results reveal two things. First, it is still a difficulty even for advanced TV learners to identify the H-falling tone. As mentioned above, this is due to the absence of a H-falling tone in VT. Second, although VT also has a mid-high level tone similar to the MC H-level tone, the L1 transfer does not benefit VT learners in learning the MC H-level tone as much as in learning the MC rising tone and the dipping tone. The mutual substitution between the H-level tone and the H-falling tone in the test indicates that the absence of a H-falling tone in VT does not merely cause difficulty in learning the H-falling tone but also interferes the acquisition of the H-level tone.

Comparing the H-level tone and the H-falling tone acoustically, both are common in the initial high pitch but different in the pitch shape. The pitch of the H-level tone remains steady from the beginning to the end. The H-falling tone initiates with a high pitch and descends to a low-pitch end with a steep slope. Given the same syllable position, the H-falling tone is higher in the starting pitch and usually shorter in the duration than the H-level tone. Perceptually, though the H-falling tone is slightly shorter and more stressed than the H-level tone, MC native speakers distinguish the two tones based on the pitch shape instead of the pitch height, syllable duration or stress. Most native speakers regard these two tones with the same high pitch without noticing the difference.

However, it seems not the same to VT learners. It is known that L2 learners tend to adopt similar features used in their L1 to interpret the sounds of L2. As a high falling pitch is not a distinctive feature used in VT, learners might adopt other acoustic cues to discriminate the H-falling tone from other tones. Hence, acoustic analysis of the stimuli and learners’ error tokens involving H-level and H-falling tones were further examined to explore this issue.

4.2 Analysis of stimuli containing H-level and H-falling tones with error tokens

The stimuli with H-level and H-falling tones were analysed via the software ‘Praat’ to collect information about the starting pitch height, tone duration, and the pitch descending slope of the two target tones. The results are given in Table 2 and 3.

Table 2: Acoustic properties of the H-level tone
Table 3: Acoustic properties of the H-falling tone

<table>
<thead>
<tr>
<th>Syllable position</th>
<th>Initial pitch height (Hz)</th>
<th>Duration (ms)</th>
<th>Falling slope (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolation</td>
<td>321–369</td>
<td>312–474</td>
<td>119–199</td>
</tr>
</tbody>
</table>

Table 4: Duration threshold of the H-level tone and the H-falling tone in different syllable position

<table>
<thead>
<tr>
<th>Tone</th>
<th>Isolation</th>
<th>Word-initial</th>
<th>Word-final</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-level</td>
<td>&gt; 500 ms</td>
<td>&gt; 400 ms</td>
<td>--</td>
</tr>
<tr>
<td>High-falling</td>
<td>&lt; 400 ms</td>
<td>&lt; 350 ms</td>
<td>&lt; 350 ms</td>
</tr>
</tbody>
</table>

Table 5: Word with the H-level tone in isolation and error tokens

<table>
<thead>
<tr>
<th>Target word</th>
<th>Error token</th>
<th>Initial pitch height (Hz)</th>
<th>Duration (ms)</th>
<th>Slope (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tī</td>
<td>1</td>
<td>316</td>
<td>490</td>
<td>17</td>
</tr>
<tr>
<td>gāo</td>
<td>0</td>
<td>293</td>
<td>606</td>
<td>15</td>
</tr>
<tr>
<td>jiāo</td>
<td>0</td>
<td>265</td>
<td>633</td>
<td>7</td>
</tr>
<tr>
<td>huā</td>
<td>0</td>
<td>304</td>
<td>545</td>
<td>12</td>
</tr>
<tr>
<td>guāi</td>
<td>0</td>
<td>311</td>
<td>545</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6: Word with the H-level tone in the word-initial position and error tokens

<table>
<thead>
<tr>
<th>Target word</th>
<th>Error token</th>
<th>Initial pitch height (Hz)</th>
<th>Duration (ms)</th>
<th>Slope (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bān</td>
<td>4</td>
<td>308</td>
<td>323</td>
<td>2</td>
</tr>
<tr>
<td>xiān</td>
<td>0</td>
<td>309</td>
<td>471</td>
<td>9</td>
</tr>
<tr>
<td>túl</td>
<td>2</td>
<td>335</td>
<td>344</td>
<td>18</td>
</tr>
<tr>
<td>jiā</td>
<td>0</td>
<td>318</td>
<td>398</td>
<td>4</td>
</tr>
<tr>
<td>dī</td>
<td>3</td>
<td>319</td>
<td>384</td>
<td>1</td>
</tr>
<tr>
<td>xīn</td>
<td>0</td>
<td>340</td>
<td>479</td>
<td>4</td>
</tr>
<tr>
<td>xiān</td>
<td>2</td>
<td>318</td>
<td>467</td>
<td>16</td>
</tr>
<tr>
<td>shū</td>
<td>1</td>
<td>347</td>
<td>341</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7: Word with the H-level tone in word-final position and error tokens

<table>
<thead>
<tr>
<th>Target word</th>
<th>Error token</th>
<th>Initial pitch height (Hz)</th>
<th>Duration (ms)</th>
<th>Slope (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>jiā</td>
<td>0</td>
<td>314</td>
<td>521</td>
<td>11</td>
</tr>
<tr>
<td>huā</td>
<td>0</td>
<td>317</td>
<td>464</td>
<td>19</td>
</tr>
<tr>
<td>dān</td>
<td>0</td>
<td>306</td>
<td>512</td>
<td>13</td>
</tr>
<tr>
<td>bān</td>
<td>0</td>
<td>304</td>
<td>468</td>
<td>13</td>
</tr>
<tr>
<td>xiāng</td>
<td>2</td>
<td>307</td>
<td>413</td>
<td>20</td>
</tr>
<tr>
<td>dān</td>
<td>1</td>
<td>261</td>
<td>486</td>
<td>31</td>
</tr>
<tr>
<td>huā</td>
<td>1</td>
<td>305</td>
<td>436</td>
<td>14</td>
</tr>
<tr>
<td>jiā</td>
<td>1</td>
<td>277</td>
<td>511</td>
<td>5</td>
</tr>
</tbody>
</table>

Given a certain syllable position, the H-falling tone is higher in the initial pitch and shorter in the duration than the H-level tone. Nevertheless, both tones share some over-mapping range in the initial pitch height and in duration. As to the pitch shape MC native speakers adopt to discriminate the two tones, it is briefly represented in the last column ‘slope’ of the table, which indicates the difference of the pitch height between the start and the end. The two target tones display clearly distinct falling slope without over-mapping.

Analysis of the relation between the error tokens and the three acoustic features reveals that the occurrence of errors seems related to the duration of the target syllable. Subjects seem to adopt the duration as another phonetic cue to discriminate the two target tones besides the pitch shape: the H-level tone should be longer than some threshold while the H-falling tone should be shorter than some threshold. Based on the duration of words with the two tones given in Table 2 and 3, the duration threshold of the two tones in different syllable position is proposed in Table 4, supported by detailed analysis afterward.

First, error tokens and the acoustic properties of the H-level tone in different syllable position are given in Table 5-7.

Apply the requirement in Table 4 to the data in each table. In Table 5, the only error occurs in the target word ‘tī’ (to kick), which is also the only word with a duration less than 500ms. Concerning the H-level tone in the word-initial position shown in Table 6, all errors occur in the words with a duration less than 400ms. The only exception is the target word ‘xiān1’ which has a duration longer than 400ms. A possible alternative account is attributed to the lexical effect. The stimulus ‘xiānhuā’ (flower) which is a real word in MC.

Concerning the word-final position, there is no clear error pattern revealed yet in terms of duration due to the small amount of errors.

Next, consider the errors of the H-falling tone. Table 8-10 list the acoustic information and error tokens of words with the H-falling tone in different position.

Table 8 and 9 clearly show that all errors occur in the words with a duration more than 400ms in the isolation position and more than 350ms in the word-initial position, respectively. The two words with the longest duration in Table 9 are also the two with the most error tokens.

In Table 10, as all target words have at least 1 error token, the duration threshold of the H-falling tone in word-final position is proposed as 350ms for the shortest duration in Table 10 is 353ms. Moreover, comparison of the two ‘dān’ in Row 5 and 7 further reveals that a lexical frequency effect might also play.
to adopt the syllable duration as a phonetic cue to discriminate these two tones. However, the duration of the tone is merely a phonetic realization rather than a distinctive feature in MC, and in the oral speech, the two tones overlap to some degree in the duration. Hence, the wrong phonetic cue VT learners adopt to identify the two tones leads to the mutual substitution between these two tones.

Based on the results, there are two suggestions for teaching VT learners the MC tones. First, teachers should explain that the major distinctive feature between the H-level tone and the H-falling tone lies on the pitch slope instead of duration and train students to focus the distinction of these two tones on the pitch slope. Second, teachers should point out that syllable duration is not neither a distinctive feature in the MC tonal system nor a reliable cue to distinguish the H-level tone and the H-falling tone as the duration of these two tones overlaps to some degree.

6. REFERENCES

SEQUENCING AND TECHNOLOGY-AIDED ACTIVITIES IN THE ACQUISITION OF FOREIGN SOUNDS

Daniela L. Martino

Universidad Nacional de La Plata – ISFD 97 – ISFDyT 8
dlmartino76@gmail.com

ABSTRACT
The various difficulties regarding the learning of sounds faced by foreign language learners have been studied from a number of perspectives in an attempt to find roots and explanations to what may hinder the acquisition of L2/FL sounds. Taking into account such perspectives and in the light of the advance of technology, a number of steps and technology-aided activities appear to be key if teachers are to help students perceive and produce L2/FL sounds successfully.

This paper reviews a number of technology tools to support the teaching of pronunciation and reports an experience in training Spanish-speaking teacher trainees in the perception and production of features of connected speech.

Keywords: segmental phonetics, EFL, technology, oral proficiency.

1. INTRODUCTION

In an attempt to explain the struggles faced by students when it comes to pronunciation in a foreign language, special attention has been given to the first language and the way it might interfere with the acquisition of L2/FL sounds. Failing to articulate a certain sound has been explained as the result of two situations: either L2 learners erroneously perceive such sound [11, 13, 5, 8] or despite being aware of it, they appear to be unable to produce it [7]. Perceptual awareness of categorical distinctions would seem to take place before the ability to implement such distinctions, with time and practice as key to establishing ‘the motor control patterns needed for producing new phones in an L2’ [7].

The different perspectives that aim at explaining how pronunciation in a foreign language is acquired are essential for the development of activities that foster effective learning. The advance of technology offers an array of tools that can undoubtedly maximize students’ learning not only aiding but also accompanying them in such process.

The present paper will describe a number of stages and technology-aided activities designed in the light of the abovementioned research done on the acquisition of L2/FL pronunciation to help learners advance from controlled to automatic processing [3].

A report will be made on the preliminary results from using the approach described in the teaching of certain features of connected speech to a group of native Spanish-speaking teacher trainees in La Plata, Buenos Aires, Argentina.

2. CONTEXT

The experience hereby reported represents an instance of common practice in Language and Oral Expression II, rather than a formal piece of research, aimed at raising students’ awareness regarding their progress by confronting their perception and production of certain features of connected speech at the beginning and at the end of the module.

The group of students in this course was made up of 25 second year teacher trainees – aged between 19 and 24 years old - who have received extensive training on isolated sounds and minimal pairs.

3. THE EXPERIENCE

The abovementioned group was given a diagnostic listening test which included both general and detailed listening tasks. Students were also asked to record themselves reading a short text provided and giving a personal opinion on an issue raised in such text. Both, the listening test and the recordings, evidenced considerable difficulty in recognizing and producing stretches with either one or a combination of the following features of connected speech: more than two weak forms, catenation, different cases of assimilation (coalescent, regressive assimilation of intensity, regressive assimilation of place, etc.) and elision of sounds. Students were not given the tests back or any feedback on their recordings. The exact same test and recordings were repeated after students were trained following the steps described in this paper. The diagnostic test and recording and the post-tests are part of a routine procedure in the course used to reflect upon the effectiveness of the tools and methods for instruction.

3.1 Stages and activities

Cece-Murcia [1] refers to four distinctive stages in the teaching of pronunciation in an attempt to go beyond students’ exposure to merely descriptive
phonetics and to apply the Communicative Approach to the teaching of pronunciation. This four-stage strategy starts with the identification of a sound or contrast which appears to be problematic and continues with the finding of a context naturally abundant of such feature. Then, a set of communication oriented tasks should be designed and finally a further set of activities to recycle the teaching point periodically. A framework composed of a set of five steps is described later including: analysis, listening discrimination, controlled practice, guided practice and communicative practice [2].

Likewise, Pennington [10] suggests a progression of tasks beginning with mechanical ones (e.g. repetition of minimal pairs) followed by contextualized tasks (e.g., repetition of key words in a listening passage); meaningful tasks (e.g., choice of correct word in a sentence or reading passage); realistic (e.g., a role-play of a situation similar to one that one may face in real life); and real (e.g., discussion of the students’ real-life situation or concerns). A combination of both approaches has been adopted in pursuit of an optimized learning process.

3.1.1 Step 1: Identification

Having identified certain problematic features, students were exposed to a text naturally rich in such features. Two tools were used to find suitable texts: Playphrase.me and Tubequizard.

Playphrase.me is a multi-modal corpus-like website that allows the search of words or phrases that have been uttered in short video clips from popular TV shows and films, providing authentic input.

Tubequizard enables teachers to search for specific words or expressions in Youtube video subtitles. Once a word or phrase is typed in, Tubequizard provides a number of suggested videos. When the text was chosen, students worked on tasks aiming at understanding its context of interaction. The objective was first for students to grasp the gist of the audio material, a task that would guide them through the text to get a general understanding.

A particular feature was then demonstrated in an authentic natural context and students were asked to recognize it.

3.1.2 Step 2: Analysis

Underhill [14] [15] contends that pronunciation teaching should ‘get out of the head and into the body’, that by making sounds ‘visible’ segmental features might become less challenging for learners to master. Likewise, Cauldwell [12] emphasizes the importance of providing students with both sight and sound shapes of words. Any sight shape version of a word – be it the transcript in ordinary spelling or the phonetic version in the dictionary – may misrepresent the sound shape in context. Having the audio immediately accessible with the text can be of invaluable help.

Based on these views, students worked with exercises designed using Sonocent’s Audio Notetaker, software that allows teachers to place sound and sight substance together.

The following is an exercise based on a short extract from the film ‘An education’, with which students had already worked in step 1.

**Figure 1**: Activity designed with Sonocent’s Audio Notetaker

The first column - the reference pane - contained questions and instructions, focusing on cases of elision of sounds, weak forms and assimilation. The second column - the text pane - provided the sight shape of the audio, which students could hear as many times as needed by clicking on the third column - the audio pane. It is interesting to note that students could use the speed control in the tool bar, which goes from 1.0 (full speed) down to 0.5. Undoubtedly, this made it easier for students to perceive different pronunciation features.

Another exercise used at this stage was designed using Aegisub, a desktop app that allows teachers to play chunks in isolation. Originally conceived for creating and modifying subtitles, it is an excellent tool to have video, audio and the sight shape of words simultaneously, with students having the possibility of playing short chunks repeatedly. Subtitles were timed, styled and built-in real time video.

A further exercise was created using TubeQuizard, a Youtube-based service that makes it possible to create quizzes for any subtitle Youtube video. The resulting quiz exposed learners to
naturally occurring instances of weak forms, assimilation, elision and catenation, which they had to identify and type in the transcript.

**Figure 2:** Activity designed with TubeQuizard

3.1.3 Step 3: Imitation

This third stage involved the repetition of models to produce a certain feature accurately and confidently. Proprioception played an important role in this stage so that students felt their articulators and could become conscious of what they were doing. The use of cellphone front cameras was key at this stage. Following Gattegno’s silent way [20], learners prepared for production by thinking about the sounds in their minds before saying them as part of the process known as ‘the Inner Workbench’ [15]. In this way, pronunciation was approached silently, with the teacher as coach instead of model, paving the way to the next stage involving controlled practice.

3.1.4 Step 4: Controlled practice

This stage involved an array of activities from less contextualised ones to those that could be equated to realistic tasks. A most controlled technique which has proved effective is Shadowing (also called shadow reading or shadow listening), a technique where learners tried to ‘speak along’ in time with an audio text, sometimes with the transcript in front of them [9]. The main objective was in line with the connection referred before between sight and sound shapes since shadowing enables learners to polish such link between the phonological realisation of words in context and their written form.

Students recorded their productions and uploaded them in the platform Soundcloud, which provides an excellent tool to share, compare and get feedback. This last feature is of particular importance since pop up comments - with peer or teacher feedback - reinforced the immediacy of auditory and visual input.

Activities at this stage also included more communicative tasks of the type of information gap activities and cued dialogues that enabled the learner to monitor for the target feature.

3.1.5 Step 5: Communication practice

In this final stage, learners engaged in less structured activities concentrating on content rather than on form, tasks that were real [10] and helped teachers to assess learner progress informally. Meaning-focused instruction led learners to the automatisation of knowledge through tasks that resembled genuine exchanges of information [6].

Discussions, debates and activities of the sort, which students carried out and recorded at the same time in the language lab, were included at this stage. The recordings were also socialised and commented on in Soundcloud.

3. RESULTS AND DISCUSSION

The results presented in Figure 3 and Figure 4 show a significant improvement in both the recognition and production of the different features of connected speech. The results reveal that after being trained following the stages and tools described, all percentages were significantly higher.

**Figure 3.** Students’ perception of features of connected speech in pre and post-tests.

**Figure 4.** Students’ production of features of connected speech in pre and post-tests.
Comments from students’ module feedback questionnaires evidenced an overall positive response from students, who reported to have particularly enjoyed the integration of technology. The most popular activities regarding perception were those involving the possibility of listening to chunks at different speeds and of working on their own rather than as a class. Using the front camera in their cellphones was reported as a revealing activity. Regarding production, listening to themselves at different stages with pop up comments added was described as the most effective awareness raising task. Students expressed their interest in using the different tools applied with tracks and videos other than those used in the course for further personal practice.

Occasional negative comments included reference to unstable Wi-Fi connection at the institute and the fact that some of the exercises used could not be downloaded.

4. CONCLUSION

The different theories that attempt to explain the acquisition of sounds - particularly of L2/FL sounds - should provide the framework for designing activities regarding segmental instruction. The application of a number of stages integrating technology - aided activities has yielded interesting results concerning effective scaffolding for learners to achieve an automatized use of features of connected speech. Technology aids learners in ways unknown before the rise of certain software, platforms and applications, most of which have not been particularly created with pedagogical ends in mind. Ubiquitous learning [4] is fostered together with learner autonomy.

The effective integration of such stages and tools enhances learners’ opportunities resulting in an improvement in students’ oral proficiency.

4. REFERENCES

USING ELECTROPALATOGRAPHY IN SECOND LANGUAGE PRONUNCIATION INSTRUCTION: A PRELIMINARY EXAMINATION OF VOICELESS GERMAN FRICATIVES

Shawn L. Nissen, Kate E. Lester, Laura Catharine Smith, Lisa D. Isaacson, Teresa R. Bell
Brigham Young University

ABSTRACT

This study examined the efficacy of using electropalatography (EPG) to create models of articulatory contact that could facilitate second language (L2) pronunciation evaluation and instruction. Specifically, the linguapalatal contact data from six native and L2 learners’ productions of the allophonic variations of the German voiceless palatal [ç] and velar fricatives [x] were examined using a relatively thin EPG sensor. Data from the native German speakers indicated distinct patterns of linguapalatal contact between the [ç] and [x] allophonic variations. Although native and L2 speakers produced [ç] in a similar manner, notable differences between the two groups of speakers were found for [x]. These data may instruct L2 learners on how to move their tongue during speech to produce German fricatives in a native-like manner.

Keywords: electropalatography, second language instruction, German, second language acquisition, pronunciation training

1. INTRODUCTION

Advances in instrumentation and computer technology, such as EPG, have allowed L2 instructors and learners the opportunity to use new and innovative methods to help individuals improve their communication. EPG or dynamic palatometry is a computer-based tracking system designed to provide real-time visual bio-feedback of how the tongue is contacting the palate during speech [2, 3]. The instructional use of this technology has increased recently, in part due to a device produced by SmartPalate International that is relatively thin and available at a lower monetary cost than previous EPG technology.

In some cases, an individual may be resistant to learning the speech sounds of a particular language in a fluent or intelligible manner, even after receiving months or years of individual tutoring or instruction. Although an L2 instructor may perceptually identify that a speaker is having difficulty producing a particular speech sound, it may be difficult to assess the basis for the problem, especially sounds that are difficult to visualize. Instructional tools like EPG may provide a method for language instructors and learners to augment traditional L2 learning approaches with another source of biofeedback, a real-time visual display of their speech movements [4].

Although very limited in scope and number, previous studies have documented positive results when EPG has been used in learning an L2. A study by Gibbon et al. (1991) found the visual feedback from EPG was useful in helping a small number of Japanese learners of English distinguish between the /r/ and /l/ phonemes during two weeks of L2 instruction. Likewise studies of native Spanish and Thai speakers learning English also found a significant reduction in accent after a relatively short period of EPG instruction [1, 6].

One of the initial steps in using EPG for L2 learning is obtaining visual models of typical linguapalatal contact from EPG data collected from native speakers. Previous research has created EPG models of palatalized Russian consonants from a small group of native speakers and L2 learners [5]. Models of native-like linguapalatal contact patterns can be used to facilitate L2 EPG pronunciation training and evaluate L2 learners’ progress in acquiring difficult to learn sound contrasts.

Thus the aim of this study was to use EPG technology to quantify the articulation patterns of native and non-native L2 speakers learning a phonetic sound contrast in German. Specifically, linguapalatal contact data from native and L2 learners’ productions of the allophonic variations of the German voiceless palatal and velar fricatives were modeled.

2. METHOD

2.1. Participants

The EPG speech production data was elicited from six L2 learners enrolled in the second semester of a university German course. The participants were native English speakers between the ages of 17 – 25 years of age. The speakers reported no known diagnosis of atypical development of their dentition or palate, and no previous history of speech or language disorders. The comparison EPG data was previously collected from a group of six native
speakers of German who identified themselves as native speakers of High German (Hochdeutsch).

2.2. EPG Sensor

Prior to the beginning of data collection, each of the speakers in the treatment group had a dental impression made by a local dentist. From this impression a 2 millimeter thin EPG sensor, similar to an orthodontic retainer, was created by SmartPalate International®. The EPG sensors were customized to fit the contours of each individual’s teeth, extending from the alveolar ridge to the back molars. The sensors contained 124 gold-plated electrodes arranged in a grid pattern across the surface of the pseudopalate. A small microprocessor, worn around the user’s neck, was used to transfer the electrode contact pattern data via a USB cord to a computer, which is then displayed on a computer screen. The EPG software allows the individual user the ability to visually compare their real-time productions to a static model or standard. The EPG device is also able to process two channels of dynamic sensor data, allowing both an instructor and L2 learner the ability to observe their lingualpalatal contact patterns simultaneously.

2.3. Stimuli

The lingualpalatal contact data from native and L2 learners’ productions of the allophonic variations of the German voiceless palatal [ç] and velar fricatives [x] (ich-laut/ach-laut) were the focus of this study. As shown in Table 1 below, the target sounds were produced in real words embedded in the carrier phrase “ich sage das Wort __” (I say the word__). Speakers produced three tokens of each word in a randomized manner.

<table>
<thead>
<tr>
<th>Target Sound</th>
<th>Preceding Vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High-Front</td>
</tr>
<tr>
<td>[ç]</td>
<td>mich</td>
</tr>
<tr>
<td></td>
<td>sich</td>
</tr>
<tr>
<td></td>
<td>dich</td>
</tr>
<tr>
<td>[x]</td>
<td>Fluch</td>
</tr>
<tr>
<td></td>
<td>such’</td>
</tr>
<tr>
<td></td>
<td>Buch</td>
</tr>
</tbody>
</table>

Table 1. List of CVC German stimuli for the elicitation of [ç] and [x].

2.4. Procedures

Prior to data collection, the fit and functioning of the sensor was confirmed, after which participants engaged in 15 minutes of conversation with the sensor in place to allow them to adapt to the physical presence of the device. Following this adaptation period, participants produced the target stimuli while seated in a sound-attenuating booth. In addition, the participants’ speech was recorded via an internal microphone located within the EPG data processing unit (yielding a low-quality recording) and an external microphone (yielding a high-quality recording). Each stimulus word was repeated three times by each speaker.

The audio recordings were used to find the onset and offset time points of each target sound, which were recorded in a text file to the nearest millisecond. The onset and offset of each fricative sound was identified by locating the time point during which there was a distinct increase or decrease in diffuse noise energy. If the beginning or end of a sound production could not be distinguished or if the participant clearly produced an incorrect word, then the data for that production would not be included in further analyses. The recorded time points were then used as input to a custom-designed MATLAB program that calculated the activation level of each electrode across a series of 20 ms time windows. Only EPG data from the analysis window with the maximal degree of contact across all of the electrodes was used in subsequent analyses.

3. RESULTS

The activation level of each EPG electrode at the point of maximal contrast was averaged across all individual words, token repetitions, and individual speakers, to create a single three-dimensional lingualpalatal contact map for each target sound. The contact patterns for [ç] productions for the native German speakers and the L2 German learners are shown in Figure 1a and 1b, respectively. Similarly, the contact data for the [x] productions preceded by a high-back vowel are shown in Figures 2a and 2b, and the [x] productions preceded by a low-back vowel are shown in Figures 3a and 3b. To illustrate contact variability across individual productions, each electrode is shaded according to the percent of overall activation, with the unshaded electrodes having the least amount of activation (0% - 25%) and the darkest shades used to represent electrodes with the highest degree of activation (75% - 100%).
**Figure 1a:** [ç] productions from native German speakers.

**Figure 2a:** [x] productions from native German speaker (high-back vowel).

**Figure 3a:** [x] productions from native German speaker (low-back vowel).

**Figure 1b:** [ç] productions from L2 learners of German.

**Figure 2b:** [x] productions from L2 learners of German (high-back vowel).

**Figure 3b:** [x] productions from L2 learners of German (low-back vowel).
4. DISCUSSION

Data from the EPG sensors collected from the native German speakers indicated distinct patterns of linguopalatal contact between the allophonic variations of the German voiceless palatal and velar fricatives. The [ç] productions were characterized by a moderate degree of lateral contact, thereby forming a relatively narrow central groove. The lateral contact did diminish in the antero-posterior dimension when reaching approximately the alveolar ridge. The pattern of contact for the L2 learners was similar to that of the native speakers, differing by a relatively small increase in anterior contact and a slight narrowing of the central groove. These differences may or may not have caused a perceptually salient difference in production.

Greater differences in the contact patterns between the native and L2 speakers were found for the [x] productions. For the native speakers the pattern of linguopalatal contact for the [x] was limited to only a few electrodes in the distal lateral and posterior corners of the EPG sensor. In contrast, the production maps for the L2 learners indicated moderate medial and anterior contact for [x] productions when preceded by a high-front vowel. Although the degree of contact was less than in the high-back vowel context, the speakers also exhibited notable medial-posterior contact when the fricative was preceded by a low-back vowel. These differences in linguopalatal contact and central groove width would likely result in a perceptual difference in sound production. However, future research is needed to empirically examine the association between the pattern of articulatory contact and the resulting speech perception.

Despite this limitation, these EPG data provide some evidence that L2 learners of German have difficulty in producing a distinct articulatory contrast between the voiceless palatal and velar fricatives in a native-like manner. Specifically, students may need to be prompted to produce the [x] with a decrease in lateral and posterior linguopalatal contact. EPG may be an effective way to provide L2 learners real-time visual bio-feedback on how to move their tongue to produce these voiceless German fricatives in a more native-like manner.

For native English students who are having difficulty learning the German voiceless palatal [ç] and velar fricative [x] contrast, EPG may provide a method for language instructors and learners to augment traditional L2 learning approaches with visual information regarding their speech movements. By attempting to match the contact patterns of the native German speakers, the L2 students may be able to improve both their production and perception of the German voiceless palatal and velar fricative contrast.

5. REFERENCES

L2 ENGLISH SPEECH RHYTHM OF JAPANESE SPEAKERS: AN ALTERNATIVE IMPLEMENTATION OF THE VARCO METRICS

Yumi Ozaki¹, Kakeru Yazawa¹, & Mariko Kondo¹²

¹ Graduate School of International Culture and Communication Studies, Waseda University
² School of International Liberal Studies, Waseda University
ozaki.yym@gmail.com, kyazawa812@gmail.com, mkondo@waseda.jp

ABSTRACT
This study compared rhythmic characteristics of English utterances spoken by three different proficiency levels of Japanese learners and native English speakers. We applied a new speaking rate normalization method to Varco metrics, which are metrics to acoustically measure speech rhythm, to investigate how the speakers’ proficiency level affects their L2 speech rhythm.

The results showed that less proficient learners had lower variability in the duration of both vowels and consonants, probably due to their incomplete acquisition of vowel reduction and frequent insertion of epenthetic vowels. Learners with higher proficiency showed more variability and were more native-like. Our implementation of the Varco metrics successfully captured the developmental sequences of L2 rhythm acquisition, which the original Varco metrics failed to show. We thus propose that our new normalization method is more robust than the existing method and is readily applicable to future L2 research and L2 education.

Keywords: Speech rhythm, second language acquisition, Varco, Japanese, English

1. INTRODUCTION
It has long been pointed out that languages have different rhythmic characteristics. According to the traditional Rhythm Class Hypothesis, languages are classified into rhythm classes such as stress-timed (e.g. English), syllable-timed (e.g. Spanish) and mora-timed (e.g. Japanese) languages. However, studies have found that such classification is not sufficient to apply to all languages [10, 13].

For more comprehensive and accurate analysis, there have been attempts to quantify speech rhythm using phonetic metrics that do not rely on language-specific phonological units. For example, Ramus et al. [11] proposed a rhythm framework called Interval Measures (IM). In IM, a phrase is divided into vocalic and consonantal intervals, where adjacent intervals of the same type are merged to make a single interval. For instance, the phrase “next Tuesday on” is segmented as /ngkstjuzdɔm/, in which the underlined segments represent vocalic intervals and the other segments represent consonantal intervals [11]. Based on the intervals, three types of rhythm metrics are calculated: %V (proportion of vocalic intervals within a given phrase), ΔV (standard deviation of vocalic intervals in the phrase), and ΔC (consonantal counterpart of ΔV). These metrics are expected to reflect rhythmic characteristics across languages. For example, stress-timed languages tend to have higher ΔV than syllable- and mora-timed languages because stressed vowels are generally longer than unstressed ones. Also, languages that allow complex syllable structure show high ΔC because the length of consonant clusters tends to vary. However, ΔV and ΔC in IM have a limitation because they are strongly affected by speaking rate [3]. Dellwo and Wagner thus introduced the Varco (‘variation coefficient’) metrics, or rate-normalized ΔV & ΔC, which can be calculated as follows:

\[ \text{VarcoΔV} = \frac{100 \times \Delta V}{\bar{V}} \]
\[ \text{VarcoΔC} = \frac{100 \times \Delta C}{\bar{C}} \]

where \( \bar{V} \) and \( \bar{C} \) are mean durations of vocalic and consonantal intervals, respectively [3].

The effect of speech rhythm on auditory impression and intelligibility of L2 speech is well-known [12], and IM and the Varco metrics have been frequently adopted to investigate it. For instance, Kawase et al. [6] examined the rhythmic characteristics of Japanese-accented English in relation to learners’ proficiency levels, using multiple rhythm metrics including ΔV, ΔC, VarcoΔV and VarcoΔC. They found lower VarcoΔV, or less variability in vocalic intervals, for Japanese speakers than for native English speakers. This can be attributed to the learners’ incomplete acquisition of English stress [7, 8, 15]. Regarding consonantal interval duration, Kawase et al. [6] did not find a significant difference in VarcoΔC between the proficiency groups, but found smaller ΔC for the inexperienced learners. The lower variability in consonantal variation may have been due to vowel epenthesis by Japanese speakers. Vowel epenthesis occurs due to differences in syllable structure between English and Japanese. English allows
syllable initial and syllable-final consonant clusters, but the mora in Japanese typically has a simple /CV/ structure. When L1 Japanese speakers pronounce English words with consonant clusters (e.g. ‘strong’ /strɔŋɡo/ (CCCVCV)), they often insert vowels between successive consonants (e.g. /sʌtɔrɔŋɡo/ (CVCVCCV/)). Thus, consonantal interval duration uniformly gets shorter and results in lower standard deviations (ΔC). The lower ΔC for inexperienced learners can be at least partially accounted for by vowel epenthesis, which tends to occur more frequently for speakers with lower proficiency [14].

While the above study [6] thoroughly investigated the rhythmic characteristics of Japanese-accented English, the metrics they used may not have sufficiently controlled speaking rate. One problem is the ‘compositionality’ of consonantal intervals. As Grabe & Low [4] suggest, whereas most vocalic intervals in English consist of a single vowel, a consonantal interval is compositional and may contain more than one segment. This means that the duration of a consonantal interval varies depending on the number of consonants it contains, regardless of speaking rate (e.g. /CCC/ is inevitably longer than /C/). VarcoΔC (2) attempts to control ΔC by using the mean duration of consonantal intervals (C̄), but C̄ itself also varies depending on the number of consonants in the interval. As will be shown later, this makes VarcoΔC less effective at detecting consonantal variability. Due to this compositionality, Grabe & Low [4] chose not to normalize consonantal interval durations, and their methodology was also adopted by Kawase et al. [6]. However, ignoring speaking rate is not a favourable solution. One reason is that speaking rate appears to influence either vocalic or consonantal duration, as will be discussed later. Also, normalizing speaking rate is crucial in L2 phonetics research because speaking rate is subject to the learners’ proficiency level.

In the present study, we used a modified version of the Varco metrics, which were designed to avoid the compositionality problem (hereafter called nVarco), to further examine the acquisition of English rhythmic structure by Japanese learners of English. We hypothesize that variability for both vocalic and consonantal intervals will increase as the learner becomes more proficient in English, which is similar to the hypothesis proposed by Kawase et al. [6].

2. METHODS

2.1 Data

Recordings of The North Wind and the Sun [5] spoken by 72 Japanese learners of English and 25 native English speakers were analyzed. The recordings were obtained from the J-AESOP corpus, an L2 English speech corpus of Japanese speakers constructed as part of the Asian English Speech eOrpus Project (AESOP) [9]. All the recordings had previously been annotated by the J-AESOP team in Praat TextGrid format [1].

Each learner’s English proficiency level was evaluated on a 9-point scale by 8 teachers of English (4 native English and 4 Japanese speakers) based on overall auditory impression of fluency and accuracy (1 = ‘very poor’, 2 = ‘poor’, 3 = ‘average’, 4 = ‘good’ and 5 = ‘very good or native-like’, with 0.5 increments). The mean score for each learner was considered to be his or her proficiency score (see Figure 1). For statistical analysis, the learners were divided into 3 proficiency groups according to their proficiency scores: high (n = 25, score ≤ 3.125), mid (n = 23, 2.615 ≤ score < 3.125), and low (n = 24, score < 2.615). Native English speakers were assigned the proficiency level of native for comparison.

2.2 Acoustic measurements & rhythm metrics

The speech data was segmented into vocalic and consonantal intervals based on the annotations. Pauses, disfluency and fillers were excluded. In our study, we adopted the inter-pausal unit (IPU) as the basic unit of analysis, which is defined as a sequence of speech sounds surrounded by silent periods that are 200 ms or longer.

Similar to (1) and (2), nVarco normalizes ΔV and ΔC (i.e. standard deviations of vocalic and consonantal intervals) in the following way:

\[ n\text{Varco}\Delta V = \frac{100 \times \Delta V}{\bar{V}_{\text{seg}}} \]
\[ n\text{Varco}\Delta C = \frac{100 \times \Delta C}{\bar{C}_{\text{seg}}} \]

where \(\bar{V}_{\text{seg}}\) and \(\bar{C}_{\text{seg}}\) are mean durations of vocalic and consonantal segments (rather than the vocalic and
consonantal intervals calculated in the original Varco) within the target IPU. This minimal modification was added to avoid the compositionality problem while maintaining the basic structure of Varco. For the sake of comparison, we also calculated VarcoΔV and VarcoΔC as well as the raw ΔV and ΔC for each IPU.

3. RESULTS

A repeated-measures MANOVA was conducted for all the rhythm metrics: raw ΔV & ΔC, VarcoΔV & VarcoΔC, and nVarcoΔV & nVarcoΔC. This analysis method was chosen to reduce the likelihood of making a type I error, and to control individual differences. When a result proved significant for a metric, we ran additional post hoc tests for the metric with Bonferroni correction.

3.1 Raw ΔV & ΔC

Figure 2 shows the mean values and standard errors of raw (i.e. unnormalized) ΔV and ΔC for different proficiency groups. According to the MANOVA, both ΔV and ΔC could highlight the group differences ($F(3,93) = 7.5498$, $p < .001$; $F(3,93) = 6.6849$, $p < .001$). Post-hoc tests found significant differences between each of the learner groups and the native speaker group for both ΔV ($p < .01$) and ΔC ($p < .05$). However, contrary to the hypothesis, the lower proficiency groups (low and mid) showed greater variability in both vocalic and consonantal intervals than the high and native groups (Figure 2). This result seems to reflect the relationship between L2 proficiency level and speaking rate. Table 1 shows the mean durations of each segment (whether vocalic or consonantal) for each proficiency group. Based on Figure 2 and Table 1, it can be inferred that slower speaking rate contributes to higher variability in both vocalic and consonantal segment duration. The result indicates the necessity to control speaking rate, not only for vocalic but also for consonantal intervals.

<table>
<thead>
<tr>
<th>low</th>
<th>mid</th>
<th>High</th>
<th>native</th>
</tr>
</thead>
<tbody>
<tr>
<td>102.76</td>
<td>97.43</td>
<td>86.26</td>
<td>72.04</td>
</tr>
</tbody>
</table>

Table 1: Mean durations of each segment (in ms) by each proficiency group.

3.2 VarcoΔV & VarcoΔC

Figure 3 shows the mean values and standard errors of VarcoΔV and VarcoΔC for the different proficiency groups. The MANOVA yielded a significant effect of proficiency group on VarcoΔV ($F(3,93) = 13.642$, $p < .001$). Differences between each of the learner groups and the native group were all statistically significant for VarcoΔV (high: $p < .01$, mid: $p < .001$, low: $p < .001$). In contrast, there was no significant effect of proficiency group on VarcoΔC. As can be seen in Figure 3, the standard errors for VarcoΔC are fairly large, suggesting that the mean duration of consonantal intervals ($\bar{C}$) used in the calculation of VarcoΔC fluctuated greatly. This fluctuation can be attributed to the sensitivity of $\bar{C}$ to the compositional nature of the consonantal intervals; non-compositional VarcoΔV showed much lower standard errors.

Figure 3: Distribution of VarcoΔV and VarcoΔC for each proficiency group. Error bars represent ±1 standard error.
3.3 nVarcoΔV & nVarcoΔC

Figure 4 shows mean values and standard errors of nVarcoΔV and nVarcoΔC for each proficiency group. There were significant effects of proficiency group on both nVarcoΔV ($F(3,93) = 15.028, p < .001$) and nVarcoΔC ($F(3,93) = 18.259, p < .001$). For nVarcoΔV, similar to VarcoΔV, the post hoc tests found significant differences between each of the learner groups and the native speaker group (high: $p < .01$, mid: $p < .001$, low: $p < .001$). For nVarcoΔC, differences between native and mid ($p < .001$) and between native and low ($p < .001$) were statistically significant. Consistent with the hypothesis, higher proficiency learners exhibited greater variability in both vocalic and consonantal intervals, suggesting that their speech rhythm became more and more native-like over L2 learning.

**Figure 4**: Distribution of nVarcoΔV and nVarcoΔC for each proficiency group. Error bars represent ±1 standard error.

4. DISCUSSION & CONCLUSIONS

This study re-examined the rhythmic characteristics of Japanese-accented English, based on speakers’ proficiency levels in L2 English. The results show that nVarco had the highest detectability and validity of all the metrics, validating our hypothesis.

The results of nVarco were consistent with our hypothesis, and showed that high proficiency learners exhibited increased and more native-like durational variability for both vocalic and consonantal intervals. nVarcoΔV seems to reflect the acquisition of English stress rhythm by Japanese learners, as found in existing studies [7, 8, 15]. Higher nVarcoΔC for high proficiency learners can be explained by the findings in previous studies that experienced Japanese learners of English produce less epenthetic vowels in L2 English [14].

The results of the original Varco metrics were similar to that of nVarco in terms of vocalic interval duration (VarcoΔV), but consonantal interval duration (VarcoΔC) revealed the methodological limitation of VarcoΔC. VarcoΔV displayed a similar tendency to that of nVarcoΔV because most of the vocalic intervals in the present study’s text consisted of a single vowel. However, VarcoΔC failed to capture the influence of L2 proficiency on rhythm characteristics. This can be attributed to the compositionality problem of consonantal intervals. The mean duration of consonantal intervals ($\bar{C}$) varied greatly, presumably because (a) the text contained consonantal intervals consisting of various numbers of consonants, and (b) $\bar{C}$ of Japanese learners further varied because of their relatively shorter and more ‘choppy’ phrasing than native speakers. For the above reasons, we argue that nVarcoΔC instead of VarcoΔC should be adopted in future L2 rhythm research. We also argue that nVarcoΔV is more robust than VarcoΔV because vocalic intervals can also be compositional in some languages (e.g. /aoiie/ ‘blue house’ in Japanese).

The results of IM (i.e. rawΔV and ΔC) in the present study contradicted our hypothesis, with native speakers having less variability in the duration of both vocalic and consonantal intervals. This strongly supports the necessity of speaking rate normalization because speakers’ proficiency level and speaking rate tend to correlate with each other in L2 speech.

In summary, the present study proposes a new implementation of the Varco metrics, nVarco, which seems to track the acquisition paths of L2 speech rhythm more successfully than the other existing metrics (IM and Varco). The nVarco metrics can quantify L2 speech rhythm more accurately and thus are expected to make contributions to future L2 rhythm research on any language sets. This study also contributes to L2 education because nVarco metrics can clearly assess the nativelikeness of learners’ L2 speech rhythm, and thus can specify which aspect of speech rhythm should be focused on in further practice.

6. ACKNOWLEDGEMENT

This study was supported by JSPS Grant-in-Aid for Scientific Research (No. 15H02729) and Waseda University Ph. D student grant for international conference.

7. REFERENCES


In Japanese phonology, a /CC/ sequence is tolerated when the first consonant is a special mora (i.e. moraic nasal or geminate consonant).
WHAT DOES THE RAPID SPREAD OF /U/-FRONTING IN AMERICAN ENGLISH HAVE TO DO WITH THE TEACHING OF FRENCH PHONETICS?

Gladys E. Saunders
University of Virginia
ges6n@virginia.edu

ABSTRACT
The rapid spread of /u/-fronting across N. America is interfering with the way American learners of French perceive, articulate, transcribe and even talk about French sounds, particularly the vowels [y] and [u]. No longer is it sufficient (or entirely accurate) for L2 French teachers to tell their students that “French [u] sounds like the vowel in the English word ‘who’ [hu:]” (because for some students the vowel in ‘who’ has shifted forward, as far forward as /i/ in some cases, and does not sound like French [u] at all, but rather more like French [y]). Here I highlight some of the phonetic research that led me to make this claim and present corroborative evidence of a different nature, namely comments made by students in a French phonetics course about their pronunciation of [u] and [y]. I also propose practical solutions to counteract the /u/-fronting interference.

Keywords. L2 French phonetics; /u/-fronting interference; student self-assessment

1. INTRODUCTION

Anyone paying attention to ongoing sound changes in the English-speaking parts of the world today is well aware that the vowel in words such as ‘food’, ‘noon’, ‘you’, ‘do’ (and of course ‘goose’) is “on the move” -- that is to say, more and more English speakers (particularly young adults and teens) are now pronouncing the vowel in these words with a greater degree of tongue-fronting than in previous decades, especially in N. America where the change is spreading rapidly (see [1] for an interesting account). The /u/-fronting phenomenon in English is also having an effect on those who study and teach French phonetics. For instance, students whose L1 English vowel inventory includes a fronted variety of /u/ will perceive, initially (until taught otherwise), French back [u] (and French front [y] as well) in terms of their L1 variant; and when they speak French this variant will also interfere with their production of [u] and [y], resulting in miscomprehension, as in [tyvabjɛ] tu vas bien (‘you are well’) for the intended [tuvabjɛ] tout va bien (‘all is well’), or vice versa.

Although there is an increasing body of research documenting /u/-fronting in North American varieties of English (see [1]), as well as a number of empirical studies reporting on the perception of French vowels by American L2 French speakers (such as [2], [3], [4]), the research seems not to have permeated the walls of the classroom. The majority of students in a French phonetics course have no clue that some of them actually pronounce the vowel in the English word ‘two’ as far forward as [ty:], and that they pronounce the French vowels [u] in tout (‘all’) and [y] in tu (‘you’) indistinguishably from their English variant in ‘two’.

2. PHONETIC RESEARCH

Comparative acoustic measurements made by Delattre [5] more than half a century ago of the American /u/ (in ‘who’) were set at (F1) 300, (F2) 900 (for an adult male speaker). In more recent studies [6, for example], greater F1 and F2 measurements have been recorded, averaging around (F1) 378/ (F2) 997 (for men), a clear indication of tongue advancement toward the front of the mouth, which is the point that I wish to make. In passing, I note that in Delattre’s comparative study, measurements for French [u] were (F1) 250 / (F2) 750, and for French [y], (F1) 250/ (F2) 1800 -- figures still quoted by French phoneticians (cf. [7]). The early acoustic values of Delattre, when taken together with the observations given in the following section, do not suggest a cause/effect relation between the ongoing /u/-fronting phenomenon in English and student pronunciation mistakes of French [u - y].

In the early 1950s, Lois Gaudin [8], a professor of French at Brooklyn College, drafted a list of the most common mistakes in French
pronunciation made by American students [a sort of precursor to ‘error analysis’ studies]. “Most mistakes, she said, are due to the fact that the student carries over into French his own English speech habits and sounds..., that he confuses sounds [...] which exist in both languages but are not identical...; [and] that he has difficulty mastering sounds [...] which do not exist in English such as [y]...”. Her only phonetic comment about [u] was one applicable to all French vowels: students fail to articulate with sufficient muscular tension (and sufficient rounding of the lips, for [y] and [u]).

It is quite unlikely that in the 1950s in the state of New York American students of French would have pronounced or perceived the vowel in the pronoun ‘vous’ [vu] (‘you’) or ‘tout’ [tu] (‘all’) as [y]. The /u/-fronting sound change would not have been prevalent on the East Coast of the US in the 1950s.

Several decades later Flege [2], in his research on new and equivalent sounds, asks whether experienced American speakers of French have more difficulty pronouncing and perceiving ‘new’ French sounds (e.g., French [y]), as opposed to similar ones (i.e., those with a phonemic equivalent in English). He concluded that similar sounds, [u] in this case, as the phoneme /u/ exists in French and English (though phonetically different), presented greater difficulties; that L2 French speakers mispronounced it more often than [y], which they tended to pronounce well (see also [9]).

Following up on Flege’s research (and challenging his conclusions about what constitutes ‘new’ vs. ‘similar’ sounds for American speakers of French, and which ones are more difficult for French-speaking Americans to perceive, French /y/ and /u/), Levy and Strange [3] designed an experiment to explore the possible role that adjacent sounds might play in the listeners’ perception of French /y/ and /u/ (i.e., the contextual variability of vowels). In his experiment, Flege [2] had cued his vowels with only the alveolar /t/, as in ‘tu’ (‘you’) and ‘tout’ (‘all). Levy and Strange [5] hypothesized that different consonants affect how one hears back and front rounded vowels. They used different consonant cells (bilabial and alveolar) and found that adjacent consonants can affect the way L2 French listeners perceive French rounded vowels.

The inexperienced listeners in their study perceived vowels as new or similar depending on the consonant surrounding the vowel. In other words, French /y/ may not be a new sound for Americans in the habit of perceiving and articulating a fronted /u/, in certain contexts, but it may constitute a new sound for those who do not.

3. STUDENT SELF-ASSESSMENT

Students in a French phonetics course were asked to self-assess their pronunciation of French, at the end of the semester, in accordance with specific guidelines they were given. They compared a recording of their first diagnostic made at the beginning of the semester with the second one made at the end of the semester. They then wrote up their results in the form of a two-page essay. The diagnostic test itself took place in the language lab where 22 students read a preselected-French text; their audio files (average length around three minutes) were collected and saved. The procedure was repeated at the end of the semester. Students listened to their recordings and typed their essays on their home computers. The text which the students recorded included 15 different occurrences of the vowel [y], e.g., une commune [yŋkɔmyn] (‘district’), multiplier [myltiˈliː] (‘multiply’); and approximately the same number of occurrences of the vowel [u], e.g., toujours [tuʒɔʁ] (‘still’), cours [kuʁ] (‘jogging’). Contrary to the research highlighted above, the student self-assessment discussed below was not a controlled experiment; nor was its primary purpose to evaluate, exclusively, students’ pronunciation of the vowels [u] and [y].

3.1. Quotes from students’ self-evaluations

As space here does not allow for a full listing of the students’ responses regarding their pronunciation of French /y/ and /u/, I cite only a few of the most instructive; and then comment on their significance for the subject at hand. Most of the students chose to write their assessment papers in French (student #2 is an exception). The citations have been translated directly from the students’ original French.

(Student 1) “In the first recording I pronounced “ou” and “u” both as [y]. It is interesting that I did the opposite of what the majority of English-speaking students who study French tend to do. I pronounced
words like course [kʊːs] (‘jogging’) and ou [u] (‘or’), with [y], instead of [u]. In my second recording my only mistake was the word cours [kœrs] (‘course’) which I pronounced like cure [kyr] (‘treatment’). It is thus clear that to pronounce [u] between consonants is difficult for me. Thus in the future I will continue to work on this”.

(Student 2) “Before taking this course I did not know that there was a difference between “ou” and “u” so I used to pronounce ‘plus’ (‘more’) like [plu] instead of [ply]. Taking a diagnostic at the beginning and end of the course is a good idea because it really lets me see what I did learn and what I am still having a hard time mastering”.

(Student 3) “Before taking this course I think that I often merged certain vowels, for example [y] and [u] and [œ] and [ɛ], among others, and I pronounced them the same. In my final recording these sounds are distinct, for example, /y/ in words like ‘commune’ [komyn] ‘district’, usine [yzin], ‘factory’. To be able to visualize the transcription of a word helps me attain the correct pronunciation; thus I think that IPA symbols are a good cognitive tool.”

(Student 4) “My second recording appears worse than the first but I think this is because I am now more conscious of my mistakes. I think I must redouble my efforts to correct my pronunciation of the vowel [y]. I see that I mispronounced the first vowel in the words usine [yzin], ‘factory’ and multiplier [mytiplje] ‘multiply’ in both recordings.”

(Student 5) “I pronounced [y] like [u] because it was a sound more familiar in English.”

(Student 6) “Before taking this course I thought that the spelling ‘u’ and ‘ou’ had the same sound and this is evident in the first recording because I pronounced them in the same way. In the second recording I think that my pronunciation of the vowel is better—not perfect, but it is clear that I paid more attention to the difference between [y] and [u].”

(Student 7) “When I made my first recording I did not know much about phonetics. My vowels were not pure; there was no difference between words which should be pronounced with [y] and those which should be pronounced with [u]. In the second recording I was more confident in my ability to pronounce the words correctly... the fronted /y/ was used more correctly than in the first recording.”

3.2. Assessing students’ self-assessments

It is interesting to note that all of the students chose to focus their attention specifically on their pronunciation of [y] (either improved or less often mistakenly pronounced) to the exclusion of [u], as if this was what was expected of them, as if French [u] did not merit a remark. A follow-up study might explain whether this omission was by inadvertence, whether the students did not actually pay attention to their pronunciation of [u], or whether it was because they felt that they had never had a problem with the pronunciation of French [u]. The guidelines may have also confused them (although these were straightforward and discussed in class).

(Student 1) is actually the only one who acknowledged that her /u/ in several words is not pronounced as a back [u] but rather as a front vowel (she says [y]). She is correct in identifying the mistake -- her vowel in cours [kœrs] (‘course’) is not a French back [u], for sure. (Student 5) says that she pronounced the front vowel [y] as a back [u] because the latter is more familiar in English, meaning, I presume, the tendency of learners to draw from their L1 inventory of sounds and thus circumvent an L2 difficulty (or putting it differently, to articulate a known sound in their L1 for a lesser known sound in the L2). Some students observed a mistake in their pronunciation of [y], but did not elaborate (see, for example, the comment of student 4, regarding her mispronunciation of the words usine [yzin] ‘factory’ and multiplier [mytiplje] ‘multiply’). Here it is likely a question of L1 interference as both words exist in English, though with different vowels in the first syllable, as opposed to French, where the grapheme ‘u’ and the phoneme [y] are constant.

One of the most useful comments, ventured by students 2, 3, 6 and 7, expresses their avowed lack of knowledge (before taking the phonetics course), of the phonetic distinction existing between the two very important French phonemes, /u/ and /y/ -- and the students responding thus were advanced French learners. It seems odd indeed that the phonetic differences between ‘tu’ (‘you’) / tout’ (‘all’), for example, or au-dessus [odsu] ‘above’/ au-dessous [odsu] (‘under’) would not have been pointed out to the students earlier-- or that they would not have queried their French teachers about the similarities and differences in pronunciation.

The findings of the students’ self-assessments corroborate Flege’s [2] and also Tranel’s [6] speculation that [y] is the main object of attention in French pronunciation exercises. Reputed to be “a very French” vowel”, [y] is one of the most
characteristic sounds of French. It is commonly considered a ‘new’ vowel for Americans and difficult to acquire. Teachers are encouraged to drill students on this sound. The back vowel [u], on the other hand, is not viewed as a ‘new’ sound; nor is it viewed as a vowel in need of special attention from teachers (but see ahead). It is generally presumed that students who articulate their American /u/ for French [u] (with added vowel tension; and without diphthongization) will produce a correct replica. What I am suggesting here is that L2 French students who grow up in an area where the /u/-fronting phenomenon is very wide spread (cf. the articulation of ‘noon’, ‘do’, ‘too’, etc. of many journalists), will be confronted with a new L1 interference, in attempting to master the correct pronunciation of French [y] and [u].

4. PRACTICAL SOLUTIONS

Practical solutions which pronunciation teachers can put to use to counteract the /u/-fronting interference described above include these two: (1) heighten student awareness of how they actually produce their L1 /u/, particularly if their English /u/ is noticeably fronted; and (2) recommend to French textbook authors, online material specialists and the like, to provide a more accurate description of the English vowel /u/ in their brief comparative remarks which accompany their presentation of French [u] -- or at the least, place accurate descriptions in their notes to the instructor.

Heightening student awareness can be done humorously: for instance, one can attempt to imitate the pronunciation of /u/-fronted English vowels heard in the speech of radio or TV journalists. Students in the class can then be asked to pronounce the words in English and to add another example. Students will listen carefully to the pronunciations of their classmates and say how many different ones they were able to identify. A student whose L1 inventory includes the fronted /u/ might be asked to describe to the class the position of his/her tongue, lips, etc. The idea is to familiarize the students with the perception and production of a sound they have never paid attention to. Exercises which entail moving the tongue frontward and backward are also useful, particularly as warmup activities at the beginning of the course. Typically, for warm-up phonetic exercises the students should be asked to stand up. Once students have been made aware of the fronted vs. back pronunciations of English /u/, ear-training exercises involving these variants and the two French high rounded vowels, [y], [u], can be added. Also, if the instructor has not previously drawn the students’ attention to the difference between a phonetic description and a phonological description, the distinction could be offered as part of the awareness activity.

Regarding the second suggestion, I cannot overemphasize the importance of providing accurate descriptive phonetic information to textbook users. The textbook is the American students’ most important source of information and students rely on its accuracy.

The most recent beginning-level French language textbook that I consulted (regarding the presentation of the French vowel [u]) states the following: “The letter combination ’ou’ sounds like the vowel sound in the English word “who”: nous [we, us], vous [you], oublier (to forget), écouter (to listen to)” -- which recalls the comment made at the outset of this paper. Remarks like these are neither very helpful to learners nor to French instructors -- and yet they persist through generations of French textbooks (mainly because textbook writers do not have the time nor often the phonetic training to keep up with phonetic research).

5. CONCLUSION

This paper looked at the phenomenon of /u/ fronting in N. American English and the subtle way it is influencing the teaching and learning of two L2 French vowels, [u] and [y]. Attention was drawn to phonetic research that assisted me in finding causes for students’ mispronunciation of the French high back vowel; and to the highly generalized and phonetically imprecise comments about English /u/, given in many French language textbooks targeting American learners of French. Finally, the paper raised a number of questions which were not answered here but will be considered in a future study.

6. REFERENCES


RESEARCH-LED TEACHING IN PHONETICS: AN EXCERISE IN RESEARCH LITERACY

Jane Setter

Department of English Language & Applied Linguistics, University of Reading, UK
j.e.setter@reading.ac.uk

ABSTRACT
Research-led teaching can take a number of different forms, including training students in research methods, exposing students to research, and engaging students in it as participants [1].

This paper reports on an exercise to engage university students in phonetic research as part of a credit-bearing module by involving them in the research as participants and using an assessed reflective exercise to improve students’ understanding of aspects of phonetic research, i.e., research design and data collection. This enabled students to evaluate not only the research methodology, but also their roles as participants and as prospective researchers, thus improving their research literacy.

Keywords: Research-led teaching.

1. INTRODUCTION
In recent years, universities in the UK and elsewhere have become increasingly interested in the concept of ‘research-led teaching’ [1, 2, 4, 5]. At the University of Reading, for example, academic principles for programme and module design include the following [internal document]:

‘Students and staff work together within a community of scholars. The curriculum engages students in research and enquiry throughout their studies. Students learn about current research in their discipline/s; engage in research discussions; are equipped to progressively develop their skills in research and enquiry; and pursue their own research and enquiry.’

This puts the onus on staff not only to expose students to relevant – and preferably recent – research and scholarship, but also to actively involve students in research in order to support and develop their academic skill set.

However, simply involving students in research is not enough. Experience from pedagogic practice in assessment and feedback has taught university lecturing staff that we need to work on students’ assessment literacy [3] in order to support their ability to truly learn and develop through the stages of their degree. The same is true of research-led teaching; scaffolding students’ experience of involvement in research will support them in progressively develop their skills. I.e., academics should be active in developing students’ research literacy.

This paper examines practice over a number of academic years in which students have taken part in phonetic research as partial assessment for the module English in the World. 10% of the assessment for this module involves students acting as research participants. However, students are not assessed on their ability to carry out the research itself, but on a guided reflective task undertaken once the data has been collected.

2. THE MODULE

2.1 Module and research study description

English in the World is an optional third year undergraduate module which runs in the spring term of the UK academic year. MA students are also invited to take this as an optional module, and it often attracts around 5-10 undergraduate Erasmus students. The module covers aspects such as the development of global Englishes, social, educational and political perspectives, and examines different varieties of English, including English-based pidgins and creoles. The aim is to widen students’ understanding of the role of English as a global language and enable them to critically evaluate that role, looking at existing research in the field.

The assessment pattern is as follows:

1. Short data analysis assignment, describing the linguistic features of a variety of English (e.g., Indian English; Jamaican English). 30%.
2. Weekly multiple choice tests delivered via the virtual learning environment Blackboard, comprising 10 questions related to module reading. 10%.
3. Research study participation. 10%
4. A two hour exam, writing essay-style answers to two questions from a list of 5 options. 50%.
The research study has been a feature of the module since its inception in academic year 2006/2007, the idea being to expose students to varieties of Global English by asking them to examine specific features of one or more variety as a research participant. We would then discuss the research study in class and find out what students had learned about the variety they had been exposed to. Students have acted as participants in a number of differently-focused research projects in the area of phonetics and phonology in global Englishes, many of which have led to conference papers and journal articles.

Prior to academic session 2009/2010, the research study was an unassessed feature of the module. Despite this, students usually did the research and useful classroom discussion followed. In academic year 2008/2009, however, it became evident that students were reluctant to take part in any non-credit-bearing activity, as only two out of a class of 25 students completed the research study materials. When asked why, they reported that they did not feel motivated to complete it as it did not contribute to the final module mark. Attempts to persuade them of the benefits of taking part in the project – detailed exposure to a variety of English they were not familiar with to help inform class discussion – fell mainly on deaf ears that year.

This led me as tutor to re-evaluate the purpose of the research study as part of the module. I decided it was essential for students’ development in research and enquiry, and that it should be incorporated into the assessment pattern. However, as well as scheduling part of a lecture to discuss the study and the variety, once completed, I decided to introduce a guided reflective activity to support their development as part of the phonetics research community.

2.2 The research study as research-led teaching

I was very keen to ensure students were getting something of pedagogical value from their participation in the research study. Although, through our ethical consent procedures, the students can withdraw their data if they wish to as long as they complete the assessment cycle, completion of the tasks which result in the data itself is not what students are assessed on. Students are assessed on their ability to evaluate the process of being a participant in research, to evaluate the research study itself, and to reflect on what they have learned by taking part in the process. Brew [1] refers to this practice as engaging students in enquiry and learning rather than simply exposing students to research.

3. PROCEDURE

3.1 Sequence of events

The materials for the research study in which students were involved in any given year was prepared prior to commencement of the module. As there have been several studies, and the studies themselves are not the focus of this paper, I will not give detailed information here. However, a list of some of the studies we have undertaken are listed below.

- Listening to other Englishes: British listeners on Hong Kong and Singapore speakers.
- The Hong Kong English accent: variation and acceptability.
- Juncture cues in Hong Kong, Singapore and British English.
- The production and perception of features of intonation and tonicity in Hong Kong and British English.
- Phonological development in the community language of Polish-English bilingual children of Polish migrants to the UK.

All studies were subject to ethical review by the University.

Students were given the research materials in Week 2 of spring term and asked to return them by Week 7. This was in order to give me and any additional members of the research team time to analyse at least a subsection of the results so they could be reported back to students in the last week of teaching (Week 11), during which time we would also discuss aspects of the variety/varieties they had been exposed to. Students were required to complete the reflective activity by the end of term, so it was not always possible to discuss their reflective comments during the last lecture.

3.2 Reflective activity

Students were required to write a reflective passage, responding to guiding questions if they wished, and to post this on Blackboard in the Discussion Board area so other students could view it and comment on it if they wished. As they were assessed for this activity, it was necessary for them to reveal their names; i.e., students were not permitted to post anonymously to this assignment discussion board.

Students were given the following instructions:

When you have completed all the tasks, write your reflective post on the following:

1. How straightforward was it to do the activities?
2. What did you find most interesting about taking part, and why?
3. What did you find most difficult?
4. What do you think the challenges are of setting up an experiment like this one?
5. If you were going to conduct this type of research, what would you change, and why?
6. Listening to the speakers doing this task, what are your impressions of them?
7. And finally: What do you think has been the most useful aspect of taking part in and reflecting on your participation in this project?

I encouraged students to view and comment on each other’s posts as part of their reflection. Comment by students happens only very rarely.

3.3 Marking scheme

The marking scheme for the research project assessment was as follows:

- Student has handed in the research materials and fully engaged with the reflective activity. 85%
- Student has handed in the research materials and engaged less fully with the activity. 55%
- Student has done the reflective activity but not handed in the research materials, or vice versa. 35%
- Student has completed neither part. 0%

Examples of a ‘fully engaged’ and ‘less engaged’ post are available for students to view. Students could withdraw their contribution to the research project after the completion of this exercise.

4. STUDENT RESPONSES

Students were told that they did not have to address the questions directly, but could post a passage incorporating responses to these general areas if they wished. The majority of them simply addressed the questions in the sequence presented in 3.2.

4.1 Question 1: How straightforward was it to do the activities?

The responses to this question varied mainly depending on the type of research activity the students were involved in that year. However, themes included the following:

- Instructions had been clear and so, even when the task was complex, they were able to work out what to do;
- Some of the terminology would not be known to participants outside the general field of Linguistics;
- If students had questions, access to the lecturer was easy and so questions could be addressed quickly.

4.2 Question 2: What did you find most interesting about taking part, and why?

I had anticipated that students might find the most interesting part to be simply the English of the stimuli speakers. However, their responses revealed that they had thought much more about the process of doing phonetic research. Examples include this one, from the study on Polish-English bilingual children:

One of the things that I found interesting about this study was the idea of ‘degree of transference’. Before the study, when the instructions were given out, I thought it was strange that we would have to assess the utterances in terms of foreign accent and felt it would be difficult to do so. However, after listening to the recordings, I realised I had been wrong and that it was easy to hear whether a child’s realisation of an utterance sounded English, or was spoken with a foreign accent. [...] I also found it interesting to listen to the recordings in the role of a researcher as it demonstrated how difficult undertaking a research project is. I remember thinking that the process for the children I had assessed was time-consuming, so it highlighted the amount of time and effort a researcher dedicates to their studies.

Here is an example from the project on listening to other Englishes:

In my opinion, the most interesting part was the rating task, because I had to pay attention to specific features of the speakers’ speech, such as the accent, the syllable length and the general prosody which usually remain unnoticed in everyday life even if they play such an important role in communication; this has been very stimulating for me, because the project gave me the opportunity of reflecting on my own speech features in order to complete the analysis, even if I am not a native speaker myself (my first language is Italian).

I was particularly pleased when students said the experience of taking part in itself had been extremely beneficial.

4.3 Question 3: What did you find most difficult?

Although students mostly said the instructions had been clear under Question 1, one of the main themes
here was difficulty following the instructions. However, most students reported that the difficulty was transitional.

The other main theme here was dealing with the speech data itself. Sometimes the recordings were not very clear, and students had problems deciding how to code the speech, for example.

4.4 Question 4: What do you think the challenges are of setting up an experiment like this one?

Although the research studies varied quite a lot in design, students reported that the following might be a challenge:

- Finding stimuli speakers;
- Finding research participants (not a problem if they are students doing the study for partial credit but, in some cases, there were participants in other countries and the issue was raised about recruiting them);
- Non-homogeneity of the research participants;
- Liaising with researchers in other countries;
- Setting up the technical aspects of the studies (e.g., recording sound files, creating slideshows, writing/adapting computer scripts);
- Dealing with children.

4.5 Question 5: If you were going to conduct this type of research, what would you change, and why?

In general, students reported that they would not change much. However, for the project on listening to Hong Kong and Singapore Englishes, where participants were asked to rate the speakers in comparison with a British English model, many students said they would have benefitted from having a recording of a British English speaker to compare the Hong Kong and Singapore samples with; their own internal voice and understanding of the phonology and pronunciation of a reference English accent was not enough.

For the Polish-English bilingual children project, students said they would attempt to obtain better recordings of the children, particularly trying to avoid having so much background noise, although they did admit it might not then be possible to record the children in a relaxed, familiar setting.

In studies where students were asked to rate speakers, some said they would prefer different mechanisms for doing so. For example, students preferred a Likert scale in comparison with one which had a ‘Strongly agree – strongly disagree’ continuum.

4.6 Question 6: Listening to the speakers doing this task, what is your impression of them?

Students tended to respond to this question in one of two ways: they would either comment on the proficiency of the speakers, or on how interesting it had been to listen to another variety of English.

In the study on the acceptability of the Hong Kong English accent, some students reported that they had felt uncomfortable rating the speakers for features such as likeability and how likely they were to have high-level jobs, saying this was not a linguistic judgement.

4.7 Question 7: What do you think has been the most useful aspect of taking part in and reflecting on your participation in this project?

This is the section under which I expected to see most evidence of the development of research literacy, and I was not disappointed. Comments included:

- It has given me an insight into the work that goes into a research project.
- It will help me structure my dissertation much more effectively.
- I have never really thought about (e.g., intonation) before in much detail and this has really helped me reflect on this aspect of English as a world language.
- I have never done anything like this before and now I have experience to draw on.
- It has challenged me to consider what aspects of speech are important in communication.

4. DISCUSSION

I have valued the opportunity to re-work the research study assignment for English in the World to be a worthwhile exercise in the development of research literacy for my students. Building in the reflective activity has been vital in this respect. Without it, the exercise could look like gratuitous use of students as research participants. With it, research participation is turned into research reflection, meeting the aims of exposing students to a variety of Global English for discussion as part of the module and supporting their development as researchers in the phonetics community. Not all students have gone on to engage in phonetic research, but they all now have an idea of what a phonetics research project might look like, and have had the opportunity to reflect on how to undertake such a study, how to be a participant, and what they need to think about when designing research studies.
5. REFERENCES


A BRIEF INTENSIVE METHOD TO HELP JAPANESE LEARNERS PERFORM ENGLISH TONICITY

Yusuke Shibata*, Masaki Taniguchi**, Young Shin Kim***

Kochi University*, Kochi University**, University College London***
yusukeshibata@kochi-u.ac.jp*, tamasaki@kochi-u.ac.jp**, young.s.kim@ucl.ac.uk ***

ABSTRACT

The teaching of English intonation is largely neglected in classrooms in Japan. This may be due to the teachers’ lack of confidence or the complexity of English intonation. The aim of this study is to examine how quickly Japanese learners can understand and learn to perform English tonicity. The experiment described here suggests that Japanese learners can be taught very rapidly to understand and perform English tonicity. The next step in this line of research will be to examine how we can train teachers to teach intonation.

Keywords: English intonation, tonicity, nuclear tones teachability, learnability, Japanese learners

1. INTRODUCTION

Pronunciation teaching in Japan faces a number of challenges. NS intelligibility of English words uttered by Japanese university students is only about 50% [1]. This is caused in large part by the lack of adequate pronunciation training provided in primary, junior and senior high schools. According to a survey conducted among Japanese university students [2], 93.1% said they had either little or no English pronunciation training in junior high school English classrooms and 93.0% said the same of senior high school.

One reason why such training is not provided is teachers’ lack of confidence [3]. When it comes to pronunciation as opposed to grammar and vocabulary teaching, there are no guidelines as to what to teach and how to teach it. Even The Course of Study for Junior High Schools: Foreign Languages (English) published by MEXT [4] does not give precise, usable guidelines on pronunciation teaching. For example it refers to “Pronouncing English sounds correctly, paying attention to such basic phonetic features as stress, intonation, and pauses,” without making clear what is meant by “correctly”.

2. IMPORTANCE OF TONICITY IN ENGLISH

Tonicity (nuclear tone placement) is an important sub-system of English intonation: “speakers of English repeatedly face three types of decision as they speak. They are: how to break the material up into chunks, what is to be accented, and what tones are to be used. These linguistic intonation systems are known respectively as tonality, tonicity and tone” (Wells [5]).

According to Wells, tonicity is probably the most important of all the systems of intonation. Tonicity affects meaning. Knowledge and practical ability with regards to tonicity can affect one’s intelligibility significantly [6].

This paper examines how quickly and easily the understanding and performance of English tonicity can be taught to Japanese learners.

3. EXPERIMENT

3.1. Subjects

The subjects were 74 Japanese learners of English from two different universities. All the subjects have Japanese as their first language and have no record of hearing impairment. They were divided into five groups, A to E. There were 17 in A, 24 in B, 16 in C, 9 in D, and 8 in E. Students in groups A, B and D were those not majoring in English, and who lacked any prior instruction on tonicity. (Group A majored in Economics, B Social Welfare, and D Education.) Those in group C had taken a course in English phonetics for a semester and received instruction on tonicity as well on tonality and tone for three ninety-minute sessions. Those in group E had taken only one ninety-minute session on the three systems of intonation prior to the experiment.

3.2 Procedure

The structure of the experiment was as follows:
3.2.1. Pre-test

The subjects were told to read aloud eight pre-test dialogues, without any instruction on tonicity or intonation. Their performances were recorded on a Zoom H6 solid state recorder with an external condenser microphone Audio-Technica AT2050.

3.2.2. Tonicity Training

The subjects were then provided with ten minutes of explanation and training on tonicity, using the pre-test dialogues. The training included not only the theory of tonicity, referring to the accenting of new information and the deaccenting of old information, but also ‘intonational gymnastics’, which illustrates the location of the nucleus by means of body movements, especially hand gestures and jumps. The training took about ten minutes.

3.2.3. Post-test

The subjects were told to read aloud eight post-test dialogues. These were different from the pre-test dialogues so as to assess the ability of the subjects to perform tonicity with new linguistic material. Their performances were recorded as before.

3.2.4. Review: Post-test Tonicity Training

The subjects were then provided with further explanation and training on tonicity, using the post-test dialogues.

3.3. Materials

The dialogues used in the pre-test and post-test are shown in 3.3.1-2. Vocabulary and grammar were carefully chosen so that the students would have no difficulty understanding the meanings of the dialogues.

3.3.1. Pre-test

Dialogue pre 1
A: Does your mother speak Korean?
B: No, but my father speaks Korean.

Dialogue pre 2
A: Do you have a car?
B: No, but I want a car.

Dialogue pre 3
A: Did you have a hot summer?
B: We had a cool summer.

Dialogue pre 4
A: Did you take the picture?
B: My sister took the picture.

Dialogue pre 5
A: Are you usually busy?
B: I am always busy.

Dialogue pre 6
A: Did you see the movie?
B: My sister saw the movie.

Dialogue pre 7
A: Is John a good swimmer?
B: He’s a very good swimmer.

Dialogue pre 8
A: May I use a pen?
B: OK, you can use this pen.

3.3.2. Post-test

Dialogue post 1
A: Is that your new bike?
B: It’s my old bike.

Dialogue post 2
A: Does your mother study science?
B: My mother teaches science.

Dialogue post 3
A: Do you have a red car?
B: I have a blue car.

Dialogue post 4
A: Can you write Chinese?
B: I can read Chinese.

Dialogue post 5
A: Does your mother have long hair?
B: She has short hair.

Dialogue post 6
A: Does your mother cook dinner?
B: My father cooks dinner.

Dialogue 7
A: Did you win a gold medal?
B: I won a silver medal.

Dialogue 8
A: Do you have two dictionaries?
B: I have three dictionaries.
4. RESULTS

The average percentages of correct answers (maximum score 100) in the pre-test and post-test for each group were as follows:

Figure 1A. Group A Pre-Test

Figure 1B. Group B Pre-Test

Figure 1C. Group C Pre-Test

Figure 1D. Group D Pre-Test

Figure 1E. Group E Pre-Test

Figure 2A. Group A Post-Test

Figure 2B. Group B Post-Test

Figure 2C. Group C Post-Test

Figure 2D. Group D Post-Test

Figure 2E. Group E Post-Test
4. DISCUSSION

In all groups, the average number of correct answers in the post-test were higher than those in the pre-test. Percentage improvements were as follows: Group A 4600%, Group B 2350%, Group C 127%, group D 627%, group E 250%. The average improvement across all groups was 305%.

As expected, Group C, which had had three 90-minute sessions of instruction and training in tonicity during the semester in which the experiment took place, had a fairly high score even in the pre-test. Group E, which had only one session, scored slightly higher in the pre-test than Groups A, B, and D.

The most remarkable finding is that the subjects in Groups A, B, and D, who had had no prior instruction on tonicity, attained almost the same level of accuracy as Groups C and E in the post-test with just ten minutes of explanation and training.

As mentioned above, the training included not only the theory of tonicity but also intonational gymnastics. The students’ understanding of tonicity seemed to be facilitated by the authors’ physical demonstration of nucleus placement in the dialogues provided. The results with Groups C and E show that the instruction of theory alone had only limited results. However, training with gymnastics produced near-perfect results.

6. CONCLUSION

English tonicity is not generally taught in Japanese schools, in spite of the fact that it is an essential element of intonation with effects on meaning. Nonetheless our experiment demonstrates that tonicity is teachable and learnable for Japanese learners of English, and that very substantial improvements can be achieved very quickly. The addition of gymnastics to theory teaching seems to have a profound effect.

The next step in this line of research will be to examine how long students can retain their knowledge and ability to perform tonicity, and to find ways in which we can train teachers to use such methods in the classroom. The authors plan to produce some tonicity teaching material for teachers of English to use at all levels, from primary school onwards. We believe it is desirable for Japanese learners of English to understand and learn to perform correct tonicity as early as possible.

7. ACKNOWLEDGEMENTS

The authors express their gratitude to all the subjects who participated in this experiment, and to Geoff Lindsey for valuable comments.

8. REFERENCES

[4] MEXT (2008) *COURSE OF STUDY for Junior High Schools Foreign Languages* (English)
for Japanese junior high school students”
Proceedings of PTLC2015.
PERCEPTION OF THE GB VOWELS /ʌ/, /æ:/, /ɒ/, AND /ɔ:/ BY SLOVENIAN SPEAKERS OF ENGLISH

Andrej Stopar

University of Ljubljana, Slovenia
andrej.stopar@ff.uni-lj.si

ABSTRACT
The paper presents a study on the perception of the General British vowels /ʌ/, /æ:/, /ɒ/, and /ɔ:/ by Slovenian students of English as a foreign language. The observed vowels are first addressed by comparing the vowel systems of General British, General American and Slovenian; then the experimental part of the study is presented. The participants’ perception of the vowels was measured pre-training and post-training. The paper discusses the overall results, the progress of the strong and the weak performers, and the most typical perception errors. The participants show significant progress in their perception of the vowels after a semester of training in phonetics and phonology. The findings are used to reflect on the impact of teaching theoretical and practical aspects of phonetics and phonology.

Keywords: perception of vowels, General British vowels, back vowels, open vowels, FL phonetics, Slovenian EFL learners

1. INTRODUCTION

1.1 Vowel perception and Slovenian EFL learners

A number of theoreticians, among them Flege [7] and Best [1], agree that the acquisition of foreign language (FL) vowels largely depends on the phonetic categories of learners’ first language (L1). Given the relative complexity of the General British (GB) vowel system, it is not surprising that FL learners often encounter difficulties in both production and perception of vowel contrasts.

According to Cruttenden [5: 109], some of the most challenging issues in the vowel space are represented by vowel groups that “are closest within the vowel space”. In the case of Slovenian learners of English, this has been experimentally attested with regard to the DRESS-, TRAP, STRUT- and NURSE-vowels: /e/, /æ/, /ʌ/, and /ɔ/. In Stopar’s experiment [11], the average result for the successful perception of these vowels was at 71.1%, with the DRESS-vowel and the TRAP-vowel exhibiting the lowest levels of successful identification. The findings were anticipated based on the differences between the GB and Standard Slovenian (StS) vowel systems (StS open-mid /e/ falls between GB DRESS-vowel and the GB TRAP-vowel) and also on the known production difficulties of Slovenian EFL speakers.

Since some studies (for instance, Escudero [6]) suggest that FL learners are quite capable of learning to perceive vowel contrasts, this paper explores whether the difficulties in the perception of vowel contrasts by Slovenian learners of English are present in other, less researched vowel combinations, and whether their perception can be improved. In particular, the perception of the GB vowels in TRAP /ʌ/, PALM /æ:/, LOT /ɒ/, and THOUGHT /ɔ:/ by Slovenian (first-year) university students English is researched. By comparing perception performance before and after a course in phonetics and phonology, the paper also considers the influence of teaching on successful perception of FL vowels.

1.2 Vowel systems compared: Slovenian, General British and General American

A comparison between the GB and StS vowel systems shows a considerable lack of vowel contrasts in the StS system (see Šuštaršič [12], Petek, Šuštaršič and Komar [9], Šuštaršič, Komar and Petek [13: 137], and Srebot-Rejč [10] for more). Slovenian EFL learners are challenged with the assimilation of thirteen GB monophthongs while mostly relying on their L1 vowel categories, which include only eight monophthongs.

The four GB vowels addressed in this study, namely, STRUT /ʌ/, PALM /æ:/, LOT /ɒ/, and THOUGHT /ɔ/, share the vowel space with three StS vowels: /a/, /ɔ/, and /o/ (in, for instance, mama ‘mother’ mož ‘husband’, oče ‘father’). Since Slovenian EFL learners are often exposed to American English (American films and series that run in Slovenia are subtitled, not dubbed), the role of American English in the perception of vowels should not be overlooked.

Three General American (GA) vowels can be identified in the relevant vowel space: /ʌ/, /a/, and /ɔ/ (see Cruttenden [5: 125-129], Wells [14: xxiv] and Ladefoged [8: 42-44]).

For an illustration of the relationship between the mentioned vowels in the three vowel systems, see Figure 1.
Examining the quadrilateral above, a number of potential challenges for Slovenian learners of English can be identified. Firstly, two different GB monophthongs, the STRUT-vowel /\a/ and the PALM-vowel /\a:/, are likely to correspond to the StS phoneme /\a/. Secondly, GB LOT /\o/ may correspond to the StS vowel /\o/ (in, for instance, oče ‘father’), whereas GB THOUGHT /\o:/ seems close to StS /\o/ (in, for instance, mož ‘husband’).

If the discussion were limited to the GB and StS vowel systems, the main perception and production issues for Slovenian speakers of English would most likely be related to the maintenance of the contrast between GB STRUT /\a/ and PALM /\a:/, with the potential complication of GB /\a:/ corresponding to GA /\e/ in words belonging to the BATH lexical set. In the back vowel region other potential pitfalls can be foreseen since GB LOT /\o/ usually corresponds to /\a/ in GA but can also correspond to /\o/ before certain fricatives and velars. Since some GA speakers lack the /\o:/ vowel and merge it with /\a/, GB /\o:/ can correspond to both GA /\a/ and /\o/ (see Cruttenden [5:129]).

As the characteristics of the three vowel systems described above are likely to interfere with the perception of vowel contrasts by Slovenian EFL learners, this study aims to determine whether this is indeed the case and to measure how successful EFL learners are in their perception of the observed vowels. Moreover, the paper discusses the extent to which the participants’ perception abilities can be improved based on a semester-long course in English phonetics and phonology, and analyses the errors that they tend to make.

2. THE STUDY

2.1 Participants and their background

The study involved the entire group of first-year students of English at the University of Ljubljana. To enroll in the Department of English, they sat the national secondary school-leaving examination, which includes a CEFR B2-level test (see [4]) in EFL.

In their first semester at the Department, the students are required to take the course English Phonetics and Phonology, which consists of fifteen 90-minute lectures and fifteen 90-minute practice classes. The course covers general phonetics while focusing on the main features of the English sound system in terms of articulatory features and auditory perception. Students are also taught phonemic and some aspects of phonetic transcription.

The study consisted of the mentioned training and two separate research stages: a pre-training perception test and a post-training perception test. One-hundred and twenty-three students completed the pre-training test, and 93 students the post-training test. Based on a questionnaire accompanying the tests, 11 participants were later excluded from the study as their L1 is not Slovenian. The results presented herein thus include the answers of 82 speakers of Slovenian who completed both tests. At the end of the experiment, their ages were between 18 and 24, with an average of 19.2 (σ=1.2, μ1/2=19).

Most participants were female (73.2%).

2.2 Instruments and procedure

The first experimental stage of the study was carried out once the students were familiarized with the GB phonetic symbols. They were asked to listen to the recordings of 17 monosyllabic words and to identify the vowels in these words. Before completing the task, the four relevant symbols were revised and the participants were also warned that not all of the words in the experiment were real words. The questionnaire also provided the IPA symbols for the targeted vowels, as well as a sample word for each vowel (Wells’s lexical sets were used for this purpose). The participants were asked to fill the gaps in the listed words with the perceived vowel (for instance, ‘Word 13: /\_t/’).

The stimuli (read by a native speaker) were as follows:

- the STRUT-vowel: /\a\!/’, /\a\!/’, /\k\r\a\p/’, and /\n\a\m/’;

**Figure 1:** The researched GB monophthongs (black) contrasted with the GA (grey) and StS vowels (white) in the same vowel space.
the PALM-vowel: /kʌːt/, /nɑːm/, /faːm/, and /krɑːk/;
the LOT-vowel: /kɒːg/, /nɒm/, /fɒt/, /kɒp/, and /bɒt/;
the THOUGHT-vowel: /kɔːt/, /kʌp/, /nɑːg/, and /fɑːm/.

The list contains 17 words, eight of which are nonsense words. Nine of the stimuli end in a voiceless consonant, and eight in a voiced one.

The second stage of the experiment was conducted four months later, at the end of the semester. The participants listened to the same recordings played in a different order. After the listening task, the respondents were also asked to provide details on their age, gender and mother tongue.

The data were analysed using Microsoft Excel. The questionnaires were marked and returned to students as a form of feedback on their progress.

In between the two experimental stages, the participants attended lectures and practical classes in English phonetics and phonology. The theoretical content relevant to the present study covered the GB sound system, the articulatory features and the distribution of GB vowels, as well as various contrastive aspects of GB and StS phonetics/phonology. Simultaneously, the practical exercises aimed directly or indirectly at vowel perception included: transcription practice (especially tasks based on dictation of individual words); listen-and-repeat exercises containing sentences centred on vowel contrasts (for instance, the recordings from Collins et al. [2] or Collins, Šuštaršič and Komar [3]); transcription of minimal pairs (dictated and written words); allophonic transcription of vowel length (pre-fortis clipping); and production exercises.

2.3 Research questions

RQ1: How successfully do Slovenian speakers of B2-level EFL perceive the GB vowels STRUT /ʌ/, PALM /ɑː/, LOT /ɒ/, and THOUGHT /ɔː/?

RQ2: Are there noticeable differences between the weak and the strong performers in the two stages of the experiment?

RQ3: What are the most typical perception errors?

2.4 Results

2.4.1 Overall results

The overall pre-training performance of the participants (RQ1) was at 84%, while the average post-training result was at 95%. Figure 2 presents the results in more detail.

The light grey line in Figure 2 represents the perception performance in the final, post-training stage of the experiment. The results of the weaker students (see the left part of the figure) were between 65% and 94%, while the majority – 54% of the participants – correctly identified all the vowels tested by the stimuli. In contrast, the shaded part of the figure represents the pre-training results with individual participants achieving between 41% and 100% correct answers. A simultaneous analysis of the two components of the figure illustrates that most participants (93%) made headway in their ability to discriminate between the four vowels. Only six students (7%) performed worse at the end of the semester than at the beginning of the semester; however, in all six cases the result was only marginally different (one wrong word more).

2.4.2 Perception of individual vowels

The results pertaining to individual vowels (RQ1) are presented in Figure 3.

The pre-training results in Figure 3 show that the LOT-vowel is the most difficult to perceive (79% correct). The STRUT-vowel and the PALM-vowel are perceived with similar success (81% and 84%, respectively), while the THOUGHT-vowel is the least challenging with a success rate of 91%.
The post-training performance demonstrates improvement in the perception of all four vowels. The progress of the students is most evident in the case of LOT /ɒ/ (94% correct). STRUT /ʌ/ and PALM /ɑ:/ exhibit an equal amount of improvement (now at 94% and 95%, respectively), while THOUGHT /ɔ:/ was perceived correctly in 97% of instances.

2.4.3 Strong and weak performers

Based on their performance in the first stage of the experiment, the participants were divided into the groups of strong and weak performers (RQ3).

At the end of the semester, the strongest performers retained their initial result (16 or 17 words correct, which represents 94% or 100%). The weaker performers, on the other hand, significantly improved their perception ability: typically, it went from 11 or 12 correct answers (65% or 71%) to 16 correct answers (94%). It is noteworthy that some students improved their perception even more substantially, for instance, from 59% correct to 100% correct.

2.4.4 Common perception errors

Table 1 presents the most frequent substitute vowels that the participants in the pre-training experiment chose instead of the ones that were targeted in the experiment (RQ3).

Table 1: The most frequent errors in the pre-training stage.

<table>
<thead>
<tr>
<th>GB vowel target</th>
<th>Correctly perceived (%)</th>
<th>Main substitute</th>
<th>Main substitute (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɔ/</td>
<td>81</td>
<td>/ɑ:/</td>
<td>12</td>
</tr>
<tr>
<td>/ɑ:/</td>
<td>84</td>
<td>/ɑ/</td>
<td>14</td>
</tr>
<tr>
<td>/ɒ/</td>
<td>79</td>
<td>/ʃ/</td>
<td>19</td>
</tr>
<tr>
<td>/ɔ:/</td>
<td>91</td>
<td>/ɒ:/</td>
<td>6</td>
</tr>
</tbody>
</table>

The most challenging vowel in the pre-training stage was the LOT-vowel. It was most often confused with the THOUGHT-vowel (and vice versa). Similarly, the PALM-vowel was typically confused with the STRUT-vowel (and vice versa).

A closer scrutiny of the data shows that the substitution of the long PALM /ɑ:/ with the short STRUT /ʌ/ occurred most often in instances where the word ended in a voiceless consonant, namely, in the stimuli /kɾaːk/ and /kɑːt/. The effect of pre-fortis clipping can also be observed in the THOUGHT-vowel /ɔ:/ where the stimuli /kɾaːp/ and /kɑːt/ were the most difficult for the participants.

The most poorly perceived stimulus in the STRUT-vowel series was /nʌm/, while among the LOT-vowel words the most challenging one was /fbt/.

Table 2 presents the most frequent substitute vowels that the participants selected in the post-training stage of the experiment.

Table 2: The most frequent errors in the post-training stage.

<table>
<thead>
<tr>
<th>GB vowel target</th>
<th>Correctly perceived (%)</th>
<th>Main substitute</th>
<th>Main substitute (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɔ/</td>
<td>94</td>
<td>/ɑ:/</td>
<td>4</td>
</tr>
<tr>
<td>/ɑ:/</td>
<td>95</td>
<td>/ɔ/</td>
<td>4</td>
</tr>
<tr>
<td>/ɒ/</td>
<td>94</td>
<td>/ɔ:/</td>
<td>5</td>
</tr>
<tr>
<td>/ɔ:/</td>
<td>97</td>
<td>/ɒ:/</td>
<td>3</td>
</tr>
</tbody>
</table>

The percentage of erroneous answers presented in Table 2 is low. However, the results pertaining to the erroneous substitutes or the most problematic words are parallel to the ones given for the pre-training stage.

With regard to the possible interaction between the GB, GA and StS vowel systems, it should be noted that the GB LOT /ɒ/ was very rarely perceived as PALM /ɑ:/ – there were 2% of such answers in the pre-training stage and 0.2% in the post-training stage. Also, the THOUGHT-vowel /ɔ:/ was typically confused with LOT /ɒ/, and only rarely with PALM /ɑ/ (2% pre-training and 0.2% post-training).

2.4.5 Nonsense and real words

Another factor that was measured in the experiment due to its potential influence on the participants’ perception errors (RQ3) is the use of nonsense words. In the pre-training experiment, the participants correctly identified the vowels in 82% of nonsense words, while the result for real words was at 85%. In the post-training stage, the difference is slightly more pronounced, despite the improvement in the performance: 92% for nonsense words and 97% for real words. It should be noted here that since the differences in the percentages are rather small, the results may not be statistically significant.

3. DISCUSSION

The experiment presented above shows that the observed GB vowels are successfully perceived by Slovenian students of EFL. The participants’ perception rate at the beginning of their tertiary studies at the English Department was high (on average, they recognized 84% of the vowels). The result had substantially improved in only four months, by the end of their first semester at the department (95% correct). We believe that the participants’ progress shows that vowel perception
can be improved in a relatively short time when supported by continuous training.

The pre-training stage of the experiment provided an insight into the perception ability of the students who had just arrived at the English Department. Their success in recognizing the vowel contrasts varied, with the lowest achievement at 41%. This variation was much less pronounced at the end of the semester, when the lowest result was at 65%. Comparing the results of the two rounds of perception tests, it is especially important that the strongest participants were able to maintain their result while the weaker students made significant headway.

With regard to the individual vowels, the post-training numbers indicate improvement across the board. The pre-training percentages were between 79 and 84 percent, whereas the post-training ones were noticeably higher, between 94 and 97 percent. The finding suggests that this section of the GB vowel space is much less challenging for Slovenian EFL students than the front open to close-mid section of the quadrilateral, which was discussed in Stopar [11].

While the study has revealed that the perception of the researched vowel space is not overly problematic for Slovenian EFL students, it has also identified a number of factors that influence the perception of GB STRUT /ɑ/, PALM /ɔː/, LOT /ə/, and THOUGHT /ʌ/ by Slovenian learners.

Firstly, the frequent substitution of long vowels with short ones (and vice versa) would suggest that vowel length is not the determining factor in vowel perception. However, a closer inspection of the participants’ choices for individual stimuli also reveals that a long vowel is much more readily substituted with a short one if the stimulus ends in a voiceless consonant. This justifies the decision to focus some of the phonetics training on allophonic features such as pre-fortis clipping.

Secondly, the overview of vowel systems in GB, GA and StS led us to predict that GA may play a role in the perception of GB vowels, especially the THOUGHT-vowel /ʌ/. At first glance this assumption was not borne out since the vowel was perceived exceptionally well by Slovenian EFL students. However, the influence of GA in this case can still be assumed if some additional factors are considered. Namely, the GA substitution of the GB THOUGHT-vowel /ɔ/ with GA vowels /ə/ or /ʌ/ may be confusing to our participants’ ears. It is likely that the GB stimulus /fɔt/ is interpreted as GA /fət/ (‘fought’) and then erroneously matched with the seemingly corresponding GB lexical set, the THOUGHT-vowel. This explains why the non-se word /fɔt/ was the most challenging LOT-word in the experiment – in more than 31% of the answers, it was categorized among the THOUGHT-vowel words. The finding illustrates that even if an EFL pronunciation course is centred on one variety of English, it cannot afford to ignore any other varieties that students might bring into the classroom.

The analysis of the stimulus /fɔt/ is also significant from the perspective of nonsense words as it demonstrates that the participants always aim to make the stimuli meaningful and that normally perform better when they recognize the words. In the case of /fɔt/, it is also plausible to assume that the respondents looked for a recognizable word that they recognize and therefore interpreted the nonsense GB word /fɔt/ as the existing word /fɔt/ (‘fought’). The effect of nonsense words has an important teaching implication since it stresses their usefulness in vowel perception tasks.

4. CONCLUSION

The comparison of the pre-training and post-training results indicates that explicit theoretical and practical instruction in English phonetics and phonology contributes to the successful perception of vowel contrasts in EFL.

While the study shows that the observed GB vowels do not represent a particularly problematic part of the vowel space for Slovenian learners of English, it also illustrates that the perception abilities acquired by the participants after their secondary schooling can be further improved by means of a specialist, tertiary-level course in phonetics and phonology. Naturally, such a course should support the existing perception ability in strong students, while simultaneously improving the performance of the weaker ones. And finally, the study programme should also take into account the different varieties of English that the students are exposed to, since they represent the starting point for their successful recognition of FL sounds.

5. REFERENCES


Online Video Assessment of Clinical Phonetic Transcription Skills in Speech and Language Pathology

Claire Timmins
University of Strathclyde
claire.timmins@strath.ac.uk

ABSTRACT
An online clinical phonetic transcription exam was designed as part of a redeveloped second year exam of clinical skills within a Speech and Language Pathology degree. This paper describes the design, implementation and evaluation of this pilot phonetic assessment.

Results show that students responded well to this method of assessment, indicating high levels of preference for the individually viewed video, and also high levels of enjoyment. These findings suggest that institutional Moodle-supported VLEs are appropriate for hosting and providing appropriate assessments to replicate clinical settings.

Keywords: Clinical phonetics, transcription, online, video.

1. INTRODUCTION
Online teaching and assessment is an increasingly desirable method for enhancing student learning within Higher Education [12]. The online learning environment can provide content such as: learning activities, resources, supports and assessment [13]. Of the activities delivered via online learning platforms, assessment has been a large focus of development and investigation. The main assessment types delivered via this medium include: written assignments, online discussions, fieldwork and tests [6]. A less frequent online assessment type is the web-based video assessment of practical skills.

Within health sciences the assessment of practical skills is often delivered under the term OSCE (Objective structured clinical examination). OSCEs have become widespread in the field of undergraduate medical education since the late 1970s [4]. They are considered highly reliable and provide equity in assessment of practical clinical skills [11]. Clinical tasks for Speech and Language Pathology (SLP) students can include taking case notes, discussing client’s needs with a parent, scoring and performing standardised assessments. These assessments have many benefits but have also been noted to require a lot of preparation, large numbers of staff and different venues [16]. In response to these criticisms, tasks have been delivered using online methods [5]. It is argued that online practical assessment engages students with high quality and consistent material which is both independent and individual [5], whilst addressing previous criticisms.

2. ONLINE PHONETICS ASSESSMENT
Competency in live phonetic transcription of speech is a core skill for Speech and Language Therapists (SLTs) [14]. Teaching of phonetic theory and transcription skills are increasingly being delivered online [9, 10, 15, 18]. However, assessment of transcription skills within SLP programmes in the UK are traditionally assessed in a face-to-face environment with students transcribing, on paper, single words (often nonsense segments) produced by their tutor. This technique is far removed from the face-to-face speech assessments experienced by SLTs in a clinical context but individual face-to-face sessions would add unnecessary strain on staff resources.

To address this problem, it is suggested that students could be assessed on their skills of transcribing clinical data from a real case of a child with an unspecified speech disorder via video presented through a VLE [7] as part of a set of OSCE tasks.

3. AIMS & METHODOLOGY
The first aim of this project was to devise a robust, reliable, timed video assessment using a Moodle lesson tool via the institutional VLE. The second aim was to evaluate the feasibility of an online examination replacing the more traditional face-to-face examination.

3.1 Context and subjects
The project was devised as part of the newly implemented OSCE within the Personal Development and Professional Practice Module in second year of a BSc Speech and Language Pathology course. The author was asked to advise on a method for assessing students’ ability to analyse a speech sample in real-time with each student viewing the same subject. This paper will initially present on the pilot run of this method but will also reflect on the subsequent examination.
The subjects involved in this project included 22 second year students. All students completed the same tasks, as this pilot was to be preparation for their final exam. No comparison groups (e.g. one group using the online examination and another using an off-line version) could be used as all students required the practice. Each student completed the online examination and a follow-up evaluation survey. In addition, 4 staff members completed the same pilot examination and the same follow-up survey.

3.2 Designing and implementing the pilot assessment

Of the many tools available via Moodle, the lesson tool offered the ability to present a progression of steps that would allow students to do the following:

- Read through instructions
- Check example audio/video
- View embedded sample client video
- Move from one video to another
- Link to Qualtrics survey

For this pilot examination, two videos were recorded of typically developing children (aged 2 years and 15 years) by members of the SLT teaching team. In each video, the children completed the 10 word Diagnostic Screening Test from the DEAP assessment [3]. In this task, clients view and name a series of pictures representing mono- and multi-syllabic target words (e.g. ‘watch’, ‘fishing’, ‘helicopter’, ‘bridge’) and the facilitating SLT completes a phonetic transcription. The time taken for the SLT to complete the transcription provides a realistic pause during which the student can complete their own transcriptions. In the final examination, a video of the same picture-naming task will be provided of a real clinical paediatric case (supplied by colleagues within the NHS) of a child displaying speech errors considered to be atypical, allowing for assessment of student skills at identifying speech errors both developmental and atypical.

The Moodle Lesson tool was used to lead students through a series of steps. These are presented in Figure 1 below. The pre-exam check video was embedded from YouTube, the assessment videos were integrated using the institutional EStream system.

![Figure 1: Flowchart of the Moodle lesson pathway](image)

3.3 The evaluation questionnaire

Qualtrics was used to create an evaluation questionnaire to be presented to students and staff at the end of the pilot examination. Adding it to the end of the task would ensure completion, particularly important as online questionnaires often tend to be low [2]. This would ensure feedback from the intended audience (the students) and staff (who could comment on the suitability for the learning outcomes).

The questionnaire included questions to reflect three areas: the nature of the task, the instructions, and the environment (question details can be found in section 4.1). The questions were presented on a Likert scale with 7 responses (from strongly agree to strongly disagree) in order to achieve optimum reliability and validity [13].

3.4 Implementation of pilot online assessment

During March (in advance of the May examination period) students attended a booked computer lab. Each student was provided with a set of headphones, an IPA and ExtIPA chart, and asked to log into their individual PC. They were then informed that they were to view two videos and complete a broad phonetic transcription [17] (completed on an offline form) for each video in real-time. Once finished the students were asked to complete the feedback survey about their examination experience. Students were also strongly advised that they couldn’t pause or rewind the videos, in order to replicate a real-life clinical experience. Invigilators monitored the students’ behaviour during the pilot examination.

4. RESULTS

The aim of this project was to pilot an online video clinical transcriptions examination with BSc Speech and Language Pathology students. The specific aims were:

1. Design a timed, online video exam as part of OSCE exam in order to remove need for
individual examinations, and provide a clinical task to all students at the same time.

2. Evaluate the use of the Moodle lesson tool for online video exams

The project achieved both of these aims. A direct outcome of this pilot study was that the online video OSCE exam was used for summative assessment during the May exam period and worked very well as a result of issues raised in this pilot. The second aim was addressed by the use of both formative and summative evaluation. Through discussion with colleagues and feedback from questionnaire data, the overall suggestion is that the Moodle Lesson tool was an appropriate choice for the online video examination, with the main benefits being: can be a timed activity, allows for a pathway, and can be password protected. The immediate problems noted from the pilot examination were: some headphones not working, and video not playing.

4.1 Evaluation results

Evaluation of the project was both qualitative and quantitative. Qualitative formative feedback [2] was provided by the Learning Technology team at Strathclyde (regarding the IT support and accommodation). Beebe, Vonderwell and Boboc [1] found that time management was an influence on the use of assessment within an online platform. Their participants particularly highlighted the student demand for quick feedback. This didn’t apply to this project, but time management was certainly an issue with regard to IT services and planning.

Quantitative feedback was gathered via an online questionnaire. 21 students completed the online questionnaire and 4 members of staff. The results from the questionnaire are presented below.

The first series of questions related to the nature of the task (Figure 2, below). 95% of respondents strongly agreed that they appreciated having their own video to view (via the PC). There was overall disagreement that one video for all students would be preferred (though 5% agreed) which is a current method of assessing this task in the UK.

Traditionally an OSCE would provide a real person for practice. The students were asked if this would be preferred, and reported a high level of uncertainty (29%) but mostly agreed that a real person would have been their preference. A similar pattern of response was noted from the staff members.

Students and staff were then asked to comment on the navigation and structure of the online exam (Figure 3, below). The navigation and instructions were well received with the majority of respondents agreeing that the exam was easy to navigate and instructions were clear and logical. There was no real preference for either online instructions or paper-based (38% reported to neither agree nor disagree). Although this was a difficult assessment regardless of the medium used, 53% of the students agreed that the online exam was enjoyable.

Environmental factors evaluated related to the lab setting, timing and audio-visual quality (see Figure 4 below). Nearly 50% of students noted levels of distractions from other students but all other measures were regarded positively (timing, audio-visual quality and lab setting). The results from the staff evaluations presented a similar pattern.

Figure 2: Results from survey regarding nature of the task: % of students responding agree, neither agree nor disagree, disagree

Figure 3: Results from survey regarding navigation and structure of online exam: % of students responding agree, neither agree nor disagree, disagree

Figure 4: Results from survey regarding environment: % of students responding agree, neither agree nor disagree, disagree
The overall student experience was positive and both the evaluation results and discussions with students helped inform the design of the final May examination.

5. CONCLUSION

The use of an online video examination in place of a traditional face-to-face/group task was well-received by the students and staff. The institutional VLE provides a variety of tools which can support this form of assessment, and should be explored further.

The pilot examination informed further implementation and allowed for student practice in this new structure. Benefits from the implementation of this assessment included the following:

1. Students experiencing clinical case at the same time (reducing student collusion)
2. Saving staff time (ultimately) while replicating real-life scenarios

Additionally, this process has encouraged the design and use of online transcription practice and formative assessment (as in [10, 15, 18]).

Although the online clinical transcription exam was implemented and completed successfully, in some ways this assignment does not align with most online assessments. While the materials were presented online (using Moodle) the completion of the assessment and the marking/feedback was completed offline. However, in clinical contexts, transcription is usually completed on paper forms.

Although the implementation was not without its problems, this method of assignment was a successful method of assessing practical clinical phonetic skills within a Speech and Language Pathology programme.

6. REFERENCES


the assessment of final year medical students. 
*Education for Health* 17 (1). 17-26.


MODELLING JAPANESE SPEAKERS’ PERCEPTUAL LEARNING OF ENGLISH /iː/ AND /ɪ/ WITHIN THE L2LP FRAMEWORK

Kakeru Yazawa¹, Mariko Kondo¹,² & Paola Escudero³

¹GSICCS, ²SILS, Waseda University, ³The MARCS Institute, Western Sydney University
kyazawa812@gmail.com, mkondo@waseda.jp, paola.escudero@westernsydney.edu.au

ABSTRACT
This study concerns Japanese speakers’ perception of American English (AE) high front vowels: /iː/ and /ɪ/. Although vowel spectra are the primary perceptual cue for this contrast for native speakers, Japanese learners of English perceive /iː/ as ‘long’ and /ɪ/ as ‘short,’ being influenced by contrastive vowel length in Japanese. However, little attention has been paid to whether they will learn to use spectral information as they become more proficient in the L2.

We ran computer simulations based on the Second Language Linguistic Perception (L2LP) model to investigate whether adult Japanese learners of English are capable of adjusting their perceptual cue weighting for AE /iː/-/ɪ/ contrast. The result of the simulations, along with our pilot experiment on real learners, suggests that such learning is possible even in adulthood provided the learner has access to a large amount of native input. Implications for late language immersion will be discussed.

Keywords: computational modelling, L2LP model, high front vowels, late language immersion

1. INTRODUCTION

Adult second language (L2) learners often have difficulty in understanding native speech and making themselves understood by native speakers. Various speech perception models, such as the Speech Learning Model (SLM) [13], the Perceptual Assimilation Model (PAM) [2] and its extension to L2 [3], indicate that the cause of such difficulty resides in the learners’ linguistic experience with the phonology of their first language (L1).

Japanese speakers’ perception of high front vowels in American English (AE)¹ illustrates this point. While Japanese has only a quantity contrast between long /iː/ and short /ɪ/, AE tense /iː/ and lax /ɪ/ differ in two properties, quality and quantity [1, 17]. The SLM and PAM(-L2) predict that native Japanese speakers would hear AE tense-lax contrast (/iː/-/ɪ/) as a realization of Japanese long-short contrast (/iː/-/ɪ/) because of their similar durational properties. Empirical evidence supports this hypothesis; studies [14, 15, 19] found that Japanese learners of North American English tend to rely on duration to identify /iː/ and /ɪ/.

What remains uncertain is how this duration-oriented perception might change with L2 learning. The SLM hypothesizes that Japanese /iː/ and AE /iː/ are perceptually linked as diaphones, as are Japanese /iː/ and AE /ɪ/, and in both cases the linked properties would ultimately be intermediate between the L1 and L2 categories. This suggests that Japanese learners of English may come to use both spectral and durational cues to perceive AE /iː/-/ɪ/. In contrast, the PAM(-L2) considers the situation as two-category assimilation, in which Japanese /iː/ is perceptually assimilated into AE /iː/ and Japanese /ɪ/ assimilated into AE /ɪ/. The model predicts that perceptual learning in this context is not very likely to occur, suggesting that the learners will continue to use duration as the primary cue. However, the exact likelihood or degree of the hypothetical learning is difficult to estimate based on these models.

The Second Language Linguistic Perception (L2LP) model [8, 9, 20] employs a different approach. It aims at modelling the entire developmental process of L2 speech perception, and is associated with a computational learning framework of the Stochastic Optimality Theory (OT) [4]. The L2LP model assumes that Japanese speakers’ perception of AE /iː/-/ɪ/ belongs to what is called the SIMILAR learning scenario, in which the existing L1 categories are initially replicated, but will then be adjusted so that their boundaries match those of the L2 contrast. The strength of the L2LP model is that the overall trajectories of such a boundary shift can be simulated based on the distributions of the target L1 and L2 sounds. This enables the model to generate very specific predictions of the learning outcomes, including the learners’ relative weighting of cues.

The current study conducted L2LP-based simulations to predict whether and how adult Japanese speakers’ perception of AE /iː/-/ɪ/ changes in English immersion settings. The results of the simulations were also compared with real learners’ perceptions observed in our pilot experiment, which was conducted after the simulations. The aim is to determine if this kind of prediction can be used to estimate and assess the effect of certain curricula and teaching methods (e.g. late language immersion) even prior to actual implementation, and to provide new insights to L2 phonetic teaching and research.
2. MODELLING PERCEPTION

2.1 Tableaux and constraints

With L2LP, perceptual mapping from auditory features to phonological categories is modelled by Optimality-Theoretic constraints like (1):

(1) “A value of x on the auditory continuum f should not be mapped to the phonological category y.”

Based on the results of previous studies [11, 12], we limited the relevant auditory continua f to the first formant (F1) and duration of vowels. Examples of these constraints are “an F1 of 300 Hz is not /iː/” and “a duration of 120 ms is not /iː/.” The mapping of auditory values (x) to vowel categories (y) is determined by the ranking of these constraints. For example, Tableau 1 shows the perception of a token of Japanese /iː/, with an F1 of 300 Hz and a duration of 70 ms (cf. Table 1). Although the spectral constraints favour the perception of /iː/, they are overridden by the constraint “a duration of 70 ms is not /iː/,” which is ranked the highest and thus is the most restrictive. This reflects the superior status of duration as a cue for Japanese /iː/-/iː; a duration of 70 ms is too short for a vowel to be perceived long.

| Tableau 1: Japanese speaker’s perception of Japanese /iː/.

<table>
<thead>
<tr>
<th>[F1=300] [dur=70]</th>
<th>[F1=300] [dur=70]</th>
<th>[F1=300] [dur=70]</th>
<th>[F1=300] [dur=70]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iː/ not /iː/</td>
<td>/iː/ not /iː/</td>
<td>/iː/ not /iː/</td>
<td>/iː/ not /iː/</td>
</tr>
<tr>
<td>/iː/</td>
<td>*!</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>/i/</td>
<td>*!</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

What makes Stochastic OT different from general OT is that the strengths of the constraints are represented by their ranking values along a continuous scale (e.g. 100.0), and that randomly generated evaluation noise (e.g. with a standard deviation of 2.0) is temporarily added to the values at each time of evaluation. For example, the constraint “a duration of 70 ms is not /iː/” in Tableau 1 may have a ranking value of 110.2, but this value is slightly distorted at each evaluation (e.g. 109.8 or 113.5). This stochastic evaluation allows the model to deal with probability and variation in perception.

2.2 Lexicon-driven perceptual learning

Perceptual learning in the L2LP model equates with error-driven updating of the ranking of the constraints. When listeners detect a mismatch between the perceived vowel and the intended one based on their lexical knowledge, they attempt to optimise the perception grammar by strengthening the constraints that contributed to the misperception and by weakening the constraints that would lead to the correct perception.

For example, consider a Japanese listener’s perception of a token of AE /iː/ (Tableau 2), which has identical acoustic values to the previous example. Despite its short duration, this token should be perceived as /iː/ rather than /iː/ because an F1 of 300 Hz is too low to be /iː/. The listener’s L1-attuned perception grammar nonetheless perceived it as /iː/, which corresponds to the Japanese short /iː/ category. However, if the listener notices a mismatch between the perceived lexical item (e.g. “pick” /piː/) and the intended one (e.g. “peak” /piː:k/), marked as “✔” in Tableau 2) through the semantic context, she will adjust her perception grammar by changing the ranking of the constraints. More specifically, the ranking values of constraints on perceiving the incorrect category /iː/ are increased (“*!”) and those on perceiving the correct one /iː/ are decreased (“*”). The ranking values are changed by a small value called plasticity, which is set to gradually decrease over time. This plasticity scheme enables age-related modelling of L2 acquisition.

| Tableau 2: Japanese speaker’s misperception of AE /iː/.

<table>
<thead>
<tr>
<th>[F1=300] [dur=70]</th>
<th>[F1=300] [dur=70]</th>
<th>[F1=300] [dur=70]</th>
<th>[F1=300] [dur=70]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iː/ not /iː/</td>
<td>/iː/ not /iː/</td>
<td>/iː/ not /iː/</td>
<td>/iː/ not /iː/</td>
</tr>
<tr>
<td>/iː/</td>
<td>*!–&gt;</td>
<td>*!–&gt;</td>
<td>*!–&gt;</td>
</tr>
<tr>
<td>/i/</td>
<td>*!–&gt;</td>
<td>*!–&gt;</td>
<td>*!–&gt;</td>
</tr>
</tbody>
</table>

3. SIMULATIONS

Using the above procedures, we ran computer simulations to model the acquisition of /iː/-/iː/ in L1 Japanese, and the subsequent acquisition of /iː/-/iː/ in L2 AE. The parameter settings were identical to the previous studies [11, 12] wherever possible.

Table 1 and Figure 1 show the average F1 and duration values of Japanese /iː/-/iː/ and AE /iː/-/iː/ spoken by male native speakers of each language, adapted from [17]. We assume that the vowels’ F1 and duration values follow normal distributions that are centred on the average values along a base-2 logarithmic scale. The standard deviations were fixed at 0.2 for F1 and 0.4 for duration.

| Table 1: Average F1 and duration values of Japanese /iː/-/iː/ and AE /iː/-/iː/ (adapted from [17]).

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Duration</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Japanese
/i/ 51 ms 317 Hz
/ii/ 138 ms 295 Hz
English
/i/ 100 ms 312 Hz
/i/ 86 ms 486 Hz

Figure 1: Average F1 and duration values of Japanese /ii/-/i/ and AE /i/-/i/ and their equal likelihood lines (solid = Japanese, dotted = AE).3

We divided the F1 continuum into 21 logarithmically equal steps ranging from $\log_2 260$ to $\log_2 500$. The duration continuum was also divided into 21 steps ranging from $\log_2 50$ to $\log_2 150$. We then created 42 constraints for the vowel /ii/, such as “an F1 of $\log_2 260$ is not /i/i” and “a duration of $\log_2 150$ is not /i/i,” and analogously for /i/. The resulting 84 constraints were used to model perceptual learning of L1 Japanese /ii/-/i/. The same constraints were subsequently re-used for learning L2 AE /i/-/i/ because the L2LP model assumes that the initial state of L2 perception is a copy of the L1 perception grammar (Full Copying hypothesis).

3.1 L1 acquisition

The ranking values of all constraints were initially set to 100. The virtual learner received 1000 tokens of Japanese /ii/ and /i/ per month, drawn randomly with equal frequency from the normal distributions described above. The learner was always informed of the intended vowel category, and every time there was a mismatch between the perceived vowel and the intended one, she changed the ranking values of the target constraints by her plasticity. The plasticity was 1.0 during the first 10 months, 0.1 between 11-100 months, and 0.01 between 101-200 months. The evaluation noise was kept constant at 2.0.

Figure 2 shows the perceptual performance of the simulated learner after 200 months of learning. This was obtained by running 1000 instances of each of the 441 F1-duration pairs through the grammar and counting the number of /ii/ and /i/ responses. The duration/spectral cue weighting, calculated by logistic regression coefficients ($\beta_{spe}c$ for F1 and $\beta_{dur}$ for duration) as follows [10], was 0.875:

\[
\beta_{dur}/(\beta_{dur} + \beta_{spe})
\]

The learner thus seems to depend almost exclusively on duration in the perception of /ii/ and /i/.

Figure 2: Simulated learner’s perception of Japanese /ii/-/i/ after 200 virtual months of learning (black = /ii/, white = /i/).

3.2 L2 acquisition

Following the L2LP model’s Full Copying hypothesis, a copy of the learner’s L1 grammar was used as her initial grammar for L2 learning. The learner was immersed in a very rich AE learning environment where she received 1000 tokens of AE /i/-/i/ per month, again drawn randomly with equal frequency from the distributions described earlier. The learning processes were identical to L1 acquisition except that the plasticity was fixed at 0.01 from the beginning.

Figure 3 shows the course of the virtual learner’s perceptual development obtained in the same way as Figure 2. The duration/spectral weighting was 0.851 after 1 month of learning, 0.697 after 10 months, 0.483 after 50 months and 0.264 after 200 months. Although the learner’s initial grammar was dependent mostly on duration, her perceptual boundary gradually shifted to incorporate more spectral information as learning proceeded. During this process, the ranking values of the duration constraints, which were ranked either very high or very low after L1 learning, gradually approached 100 again. In contrast, the ranking values of F1 constraints either increased or decreased to play a major role in L2 perception. Ultimately, F1 became the dominant cue for identifying AE /i/-/i/.

Figure 3: Simulated learner’s perception of AE /i/-/i/ after 1, 10, 50 and 200 months of learning (black = /i/, white = /i/).
4. PERCEPTUAL EXPERIMENT

In order to verify the result of the simulations, we then conducted a pilot experiment on 4 Japanese learners of English (A, B, C, D, mean age = 23.3). They had received formal English language education in secondary schools in Japan (age 13-18) at the time of the experiment. A, B and C were graduate and undergraduate students enrolled in English immersion programmes at Waseda University for 8, 7 and 4 years, respectively. D had little or no daily exposure to English, and therefore serves as a direct test of the Full Copying hypothesis and represents the initial state of L2 perception.

The stimuli were 49 synthetic vowel-like sounds, with F1 ranging from 260 Hz to 500 Hz in 7 logarithmically equal steps, and duration from 50 ms to 150 ms in 7 logarithmic steps. The second formant (F2) value was set to co-vari with the F1, ranging from 2300 Hz to 1800 Hz on a logarithmic scale, to make the stimuli sound more natural. The fundamental frequency was fixed at 120 Hz. The stimuli were created using Praat’s KlettGrid [7].

The experimental design was created using Praat’s Multiple Forced Choice (MFC) format. The participants heard the stimuli in isolation over headphones and decided whether the vowel was /i:/ or /u/ by clicking the words “beat” or “bit” displayed on a computer screen. The 49 stimuli were presented randomly in 10 randomization blocks, resulting in a total of 490 responses. Participants were given a short break after each randomization block.

Figure 4 shows the result of the experiment. As expected, D appears to rely exclusively on duration to distinguish /i:/ and /u/ with a duration/spectral cue weighting of 0.982, which suggests that she used a copy of her Japanese perception grammar. C also used duration as the primary perceptual cue but to a slightly lesser extent with a cue weighting of 0.857, indicating that her reliance had not changed despite her extended exposure to English. In contrast, A and B seemed to distinguish the contrast mostly by vowel spectra. Their cue weightings were 0.028 and 0.066, respectively. Overall, the result shows close resemblance to the simulated stages in Figure 3, indicating that at least some learners can achieve native-like (i.e. spectra-based) L2 perception.

5. DISCUSSION AND CONCLUSION

The overall result of the current study confirms the L2LP model’s prediction that Japanese speakers’ perceptual boundary between AE /i:/ and /u/ will shift with L2 learning. The simulated learner initially used duration as the primary perceptual cue, which matches the perceptual behaviour of learners D and C. Then the virtual learner’s perceptual boundary gradually shifted as she received more input, and ultimately the spectral cue became more important than duration. Learners A and B’s spectra-based perception supports this prediction.

The L2LP model claims that L2 learners are optimal perceivers. According to the model, an optimal perceiver will prefer auditory dimensions that reliably differentiate the sound contrasts in the target language. In the case of Japanese speakers’ perception of AE /i:/ and /u/, they are expected to learn to prefer vowel spectra to duration because spectral information is a more reliable cue in the production environment of AE. The model proposes that optimal perception in L2 is possible because L2 learners have direct access to the distributional learning device they use during L1 acquisition (the Gradual Learning Algorithm [6]), which enables L1-like boundary adjustment and new category formation. However, this does not necessarily mean that L2 acquisition occurs as efficiently and flawlessly as L1 acquisition. Based on past research [16, 18], it is probably undeniable that there is a negative relationship between the learners’ age and their ultimate attainment of L2 pronunciation.

Importantly, the L2LP model does not reject the possibility of age-related limitations on L2 acquisition, as is reflected in the plasticity scheme in the simulations. Yet, Escudero [8] argues that rich input in the learning environment overrules cognitive plasticity. This argument seems in line with the results of our experiment. Although all of the participants were sequential rather than simultaneous bilinguals, some learners (A and B) perceived the L2 contrast in a native-like fashion. Learner C’s duration-based perception was somewhat unexpected, but it could perhaps be attributed to her relative inexperience in the immersion environment.
compared to A and B, and her perception may possibly change afterwards.

In conclusion, the predictions made by the L2LP model, supported by the perceptual experiment, seems to suggest that late language immersion can be effective for at least some learners. Although our focus was on English immersion, L2LP-based simulations are amenable to various types and amount of input, and thus can be applied to investigate other teaching/learning environments.

6. ACKNOWLEDGEMENT

This study was supported by JSPS Grant-in-Aid for Scientific Research (No. 15H02729) and Waseda University Ph. D student grant for international conference.

7. REFERENCES


1 This study focuses on American English because it is the accent used in formal English language education in Japan.
2 Stochastic OT can model how phonological categories emerge before the lexicon is established [5]. However, since category emergence is beyond the scope of this paper, we assume that the learner has already established her lexical representations prior to learning.
3 The slopes of the equal-likelihood lines can be calculated as the ratio of the F1 and duration distances between the two vowels in each language, multiplied by the squared ratio of the duration and F1 standard deviations (0.4 and 0.2, respectively). The lines go through the midpoints of the vowels in each language.
4 Learners A and C also studied in the United States for 1 year during their undergraduate program.