ENGLISH CENTERING DIPHTHONG PRODUCTION BY POLISH LEARNERS OF ENGLISH

Anna Balas
abalas@ifa.amu.edu.pl
Adam Mickiewicz University, Poznań, Poland

Abstract
The paper shows how British English centering diphthongs are adapted to the vowel space of Polish learners of English. The goal is to focus on complex vowels and the interaction of qualitative and quantitative features. Acoustic analysis revealed various processes used to overcome pronunciation difficulties: /j/ and /w/ breaking, /r/ insertion, substitutions of other vocalic qualities, changes in diphthong duration and diphthong phases duration, and changes in the rate of frequency change.

Keywords: centering diphthong, Polish, British English, Natural Phonology

1. Introduction

This paper is devoted to a differential study of centering diphthongs in triangulation: it compares English centering diphthongs produced by Polish learners of English to hypothesized Polish sounds as source models and to English centering diphthongs as target models. Centering diphthongs seem to be even more challenging for Polish learners of English than rising diphthongs as their production involves both substitution of sounds for English targets and processes employed to deal with a sequence of two vocalic qualities. Section two offers a summary of previous phonetic studies of English centering diphthongs. Section three is devoted to hypothesizing about the nature of centering diphthong production by Polish learners of English. Section four presents the methodology of the experiment and is followed by sections discussing the results and presenting conclusions.

2. Previous phonetic studies of English centering diphthongs

Cruttenden (2001) distinguishes between almost complete changes, well-established changes and recent innovations. Among almost complete changes related to diphthongs he notes that /eə/ is realized monophthongally as /eː/ and that /əʊ/ is regularly pronounced as /əʊ/ rather than as /oʊ/. Among well-established changes related to diphthongs he notes that /ɔː/ is used instead of /ʊə/ (the same observation is made by Roach et al. (2006)), especially in monosyllabic words, like *sure*, *poor*, *cure*, *moor*, and

tour. This change, however, is lexically conditioned. Certain words like monosyllabic *pure* and non-monosyllabic *curious*, *puerile*, *endure* and *secure* are less likely to be produced with /ɔː/, whereas others like *dour*, *gourd*, *lure*, *Ruhr* and *Ure* are never pronounced with /ɔː/. Also words derived from /uː/ plus a suffixal /ə/ like *doer*, *fewer*, *newer*, *two-er*, *viewer* are never pronounced with /ɔː/. Among recent innovations he cites realizing /ɪə/ and /ʊə/ as [ɪː] and [ʊː], e.g. *beer* [bɪː] and *sure* [ʃʊː]. N.B. the latter change competes with the replacement of /ʊə/ by /ɔː/.

Upton et al. (2003) discuss changes related to two diphthongs. Their transcription of the diphthong in words such as *nice* and *try* is /ʌɪ/ because they judge the starting point of the diphthong to be now in the area of the half-open, back centralized vowel /ʌ/, rather than the low front position [a]. They add that this symbol was first used for RP by MacCarthy (1978). The second diphthong they comment on is the one in *square* and *hair*. They note that a fully diphthongal pronunciation is now a feature of a marked variety of RP, even more rarely heard in a compound such as *hairpiece* than in *hair*. The primary pronunciation according to Upton et al. (2003) is now a monophthongal /ɛː/, although especially in a stressed final syllable it might be followed by an off-glide.

The diphthong formant values are also available (see Table 1) thanks to acoustic studies in the Phonetics Laboratory, Department of Linguistics, University of Manchester for the purpose of Cruttenden (2001). Only diphthongs in citation form were tested, so the results will not be comparable with the ones for diphthongs in context presented in this study.

Table 1. Formant frequencies (Hz) for RP diphthongs (in citation form) taken from Cruttenden (2001)

Diph-thongs	First component				Second component			
	F1		F2		F1		F2	
	Male	Female	Male	Female	Male	Female	Male	Female
ɪə	382	399	2096	2514	578	417	1643	1846
eə	538	691	1864	2210	655	751	1594	1883
ʊə	426	420	1028	1157	587	485	1250	1258

3. The hypothesis

This section presents the hypothesis on the nature and mechanisms of English diphthong production by Polish adult learners of English and adaptation of diphthongs to the existing vowel space. The hypothesis is a scenario from which certain predictions follow. The approach taken here is termed "Vowel Space Repopulation" to emphasize that the vowel space of Polish learners of English is restructured by complex processes forming new categories and these phenomena are not adequately captured by notions such as "transfer" and "interference". The hypothesis presents how Polish vowel categories and processes are used by English-speaking Poles and how universal and L2

categories and processes might be introduced to the interlanguage[1] phonological system of Poles learning English.

3.1. The L2 phonological system

Natural Phonology[2], a functional theory, is adopted in the present paper in order to model and explain second language acquisition phenomena in English diphthong production by Polish learners of English. The choice of the theory is dictated by the assumption that phonetic detail plays a crucial role in phonology in second language acquisition. The suitability of Natural Phonology for second language acquisition research is seen in its views on the phonetics-phonology relationship, advantages over formal phonological theories, and aims of pronunciation training.

Polish adult learners of English do not start learning English in a vacuum. It has long been suggested that L1 acts as a "sieve" filtering out speech features which are not significant in the first language phonological system (Polivanov, 1931, Trubetzkoy, 1939/69). The hypothesis of the present paper adopts this claim. A particular contribution of Natural Phonology to second language phonology research is that L2 learners are equipped with L1 categories, or to be more precise, underlying representations as specified by L1, and that L1 dynamic, preference-based, subconscious processes are used to shape sounds and sound sequences in interlanguage. The use of L2 processes, which are not used in L1, may become evident. In new contexts, universal processes, whose use is not evident in either L1 or L2, are used in second language acquisition. The use of universal processes, not used in either L1 or L2, is present when the process did not have a chance to emerge in L1 because of the lack of a specific context, its use is restricted or suppressed in L2, but L2 learners have not managed to limit the process in accordance with L2 phonology (e.g. as it often happens with Japanese learners of English who devoice final obstruents, although Japanese does not have final obstruents, so it is not a process transferred from L1). With time, more universal and L2 processes come into play, as the learner notices that L1 processes are not sufficient to represent L2 sounds. The aspect of time needed for the transfer from L1 processes to universal processes and to L2 processes is not an aspect of the present paper.

According to Natural Phonology, the task of the second language learner consists in deciphering underlying representations in L2 in the perception direction of speech

---

[1] The notion *interlanguage* is not used here exactly in the sense introduced by Selinker (1972). *Interlanguage phonology* in the present paper means a phonological system of an L2 learner who uses L1, L2 and universal processes when speaking L2.

[2] Natural Phonology was founded by Stampe (1969) and developed by Donegan (1985, 1993, 2001), Stampe (1979), and Donegan and Stampe (1979). Natural Linguistics is the name which refers to the model as modified and expanded by Dressler (1984, 1985), and Dziubalska-Kołaczyk (1990, 1995, 2001, 2002a, 2002b). The model of Natural Phonology considered in this paper is not only that of Stampe (1969), and the emphasis is on later works: Donegan (1985, 1993, 2001), Dressler (1984, 1985, 1996, 1999), Dziubalska-Kołaczyk (1990, 1995, 2001, 2002a, 2002b), and Ritt (2001).

processing and arriving at phonetic representations by means of processes used in L2 in the production direction. Vital for native-like pronunciation in L2 is finding out:

- which processes used in L1 are also used in L2
- which processes used in L1 should be limited in L2
- which processes which do not play a role in L1 should be used in L2
- which universal processes which do not emerge in L1 due to the lack of a specific context, have a limited use in L2. Then the learner has to learn to limit the process in accordance with L2 phonology.

The attempts of Polish learners of English to produce English diphthongs are hypothesized to be based on L1 categories and processes, and in new contexts on universal processes, with the goal of using L2 processes. The use of L1 categories and processes is hypothesized to be dependent on the subconsciously perceived similarity of the English target on the basis of Polish processes. The use of universal processes is, however, expected to be related to:

- dissimilation of vowel qualities within the diphthong (Donegan (1985), Liljencrants and Lindblom 1972)
- as redundant features may affect the strength with which a distinctive feature is implemented (see Stevens et al. (1986) for a discussion of phonetic evidence, and Donegan (1985) for phonological explanation), second language learners' pronunciation might also be influenced by redundant features not being examined in the experiment, as their influence has not yet been attested
- approximating the tongue height should be favored over approximating the tongue backness in the case of non-low front vowel region, as in this region F2, F3, and F4 are rather insensitive to anterior-posterior displacement of the tongue body (Stevens, 1989)
- factors determining syllabicity shift (Donegan, 1985)
  - o the relative degrees of sonority of the two segments
  - o the syllable type preference
  - o the timing system of the language

3.3. Category-related changes applied to English diphthongs by Polish learners

The hypothesis is based on the assumption that Polish does not have diphthongs and it tries to give an answer to the question about the sounds substituted for English diphthongs by Polish learners of English. Polish has vowel plus glide sequences to some extent similar to British rising diphthongs.

The mismatch which exists between Polish vowel plus glide sequences and English diphthongs is hypothesized to result in non-native formant values and non-native timing relations of diphthong targets[3]. The first target of each diphthong is hypothesized to have formant values similar to respective Polish vowel formants.

---

[3] The term *target* is used here to mean a component, constituent or part of a diphthong, following the convention set by Lehiste and Peterson 1961. Therefore the term *target* as used in this paper

This section is devoted to hypothesizing about the new vowel space of Polish learners of English, specifically about acoustic properties of English diphthongs produced by Poles. Polish does not have diphthongs, though it has vowel plus glide sequences comparable to English rising diphthongs, but not to centering diphthongs.

The diphthongs containing schwa are likely to be especially difficult for Polish learners for both qualitative and quantitative reasons. Moreover, the /ɪə/ and /eə/ diphthongs are especially likely to undergo /j/ breaking and the /ʊə/ is likely to undergo /w/ breaking. Tables 3 and 4 present a comparison of the differences between English average centering diphthong formants and average formant values of Polish vowels hypothesized to be substituted for English diphthongs. The major disadvantage of this comparison is that the data come from studies performed in different conditions and with a very small number of subjects, so the results may not be very precise, and in fact they often contradict auditory impressions. They are supposed to serve as an illustration of the hypothesis, and not as precise indicators of the difference - it is the task of the experiment presented in Section 4 to present comparable values obtained in similar conditions from native Polish and English speakers.

Table 2. Formant frequencies for RP diphthongs (in citation form) (Cruttenden, 2001) and Polish vowels (Majewski and Hollien, 1967) for male speakers

Diph-thong	Hypo-thesized Polish substitute	1st element		1st substitute		2nd element		2nd substitute	
		F1	F2	F1	F2	F1	F2	F1	F2
/ɪə/	/iɛ/	382	2096	245	2286	578	1643	543	1945
/eə/	/ɛa/	538	1864	543	1872	655	1594	729	1281
/ʊə/	/uɛ/	426	1028	332	725	587	1250	543	1945

Table 3. Formant frequencies for RP diphthongs (in citation form) (Cruttenden, 2001) and Polish vowels (Majewski and Hollien, 1967) for female speakers

Diph-thong	Hypo-thesized Polish substitute	1st element		1st substitute		2nd element		2nd substitute	
		F1	F2	F1	F2	F1	F2	F1	F2
/ɪə/	/iɛ/	399	2514	341	2543	417	1846	602	2084
/eə/	/ɛa/	691	2210	602	2084	751	1883	921	1564
/ʊə/	/uɛ/	420	1157	454	834	485	1258	602	2084

---

should not be associated with a target that the tongue never reaches, as it is the case in target undershoot.

3.4. Hypothesized production of centering diphthongs

Centering diphthongs, i.e. /ɪə, eə, ʊə/, are claimed by Sobkowiak (2004) to be difficult to pronounce for Polish learners of English for three main reasons. Firstly, during their production the tongue aims at a mid-central position in the vowel space, which is not used to articulate any Polish vowel. Secondly, learners feel that, unlike closing diphthongs, centering diphthongs do not have any counterparts in Polish. Thirdly, As auditorily sustained, centering diphthongs are prone to reductions and assimilations, which make them even less distinct.

Poles are hypothesized to begin articulating centering diphthongs at positions typical for Polish monophthongs /i, ɛ/ and /u/ and, instead of more centralized positions assimilated to the following schwa. Moreover, centering diphthongs, consisting of two vocalic qualities, act against the natural tendency of a language for a CV syllable structure. Boasting consonant clusters, Polish has many exceptions to the CV structure, but no syllable in Polish contains two vowels. As a remedy in the context where two vowels occur next to each other across a syllable boundary, there is a sequence-optimizing process inserting a glottal stop or a glide between them. Glide insertion is preferred when one of the vowels is high, as in *sytuacja* /sɨtuwatsja/ (situation), *słabeusz* /swabɛwuʂ/ (weakling), *nieinteresujące* /ɲɛjintɛrɛsujontsɛ/ (uninteresting). The hypothesis is that Polish learners transfer the process of glide insertion to English centering diphthongs /ɪə/ and /ʊə/. The articulatory result is that the tongue goes from /i/ or /u/ position up outside the vowel space producing a clear glide /j/ or /w/ and then goes down again. It is also hypothesized that Polish learners of English should have problems with the final element of a centering diphthong. The final element of a centering diphthong should be articulated in a mid-central position, but this position is not used in Polish.

The /ɪə/ diphthong is hypothesized to be pronounced with the help of a sequence-optimizing process. It is expected that either a glide will be inserted or the first element of the diphthong will change into /j/, thanks to which a consonant-vowel sequence, favored over a hiatus, will be formed. As far as substitutions are concerned, there are a few possibilities. The first element of the diphthong might be realized either as Polish /i/ or /ɨ/ or changed to /j/. The /ə/ element is likely to be changed to Polish /ɛ, ɨ/ or /a/ (Bogacka [Balas] et al., 2005). The acoustic studies of English diphthongs (Cruttenden, 2001) showed the following formant values for the diphthong /ɪə/: the first element's F1 was 382 Hz for males and 399 Hz for females, the first element's F2 was 2096 Hz for males and 2514 Hz for females, the second element's F1 was 578 Hz for males and 417 Hz for females, and the second element's F2 was 1643 Hz for males and 1846 Hz for females. The formant values for Polish /i/ were as follows: F1 was 261 Hz for males and 341 Hz for females, F2 was 2280 Hz for males and 2543 Hz for females. If /ɨ/ were substituted the results would be slightly closer, especially for males: F1 was 389 Hz for males and 450 Hz for females, and F2 was 1984 for males and 2254 for females.

The /eə/ diphthong is hypothesized to be pronounced either as /ɛa/, /ɛjɛ/, /ɛja/. The first option, without an inserted glide, is hypothesized to be /ɛa/ and not /ɛɛ/. Polish does

not have long vowel or geminate vowels, therefore it is not expected that Polish learners of English substitute them for English diphthongs. It is, however, expected that they will notice the difference in vowel quality, and that they will express this difference by means available to them, like for instance changing the vowel quality as in /ɛa/. The options with an inserted glide /ɛjɛ/ and /ɛja/ are also hypothesized as it is expected that Poles will somehow want to break up a hiatus. When a glide is introduced between the two vowels, the substitution of /ɛ/ for schwa becomes possible additionally to substitution of /a/ hypothesized for the hiatus, because the two /ɛ/s do not form a long vowel or a geminate any more. According to acoustic studies (Cruttenden, 2001) the English diphthong /eə/ has the following values. The first element of the diphthong has F1 of 538 Hz for males and 691 Hz for females, and F2 of 1864 Hz for males and 2210 Hz for females. In the second element of the diphthong F1 is 655 Hz for males and 751 Hz for females, whereas F2 is 1594 Hz for males and 1883 for females. In the /ɛ/ vowel in Polish F1 is 537 Hz for males, 602 Hz for females, and F2 is 1862 Hz for males and 2084 for females. The Polish /ɛ/ seems to be a good substitute for the first part of the English diphthong, especially for males. If Polish /a/ was substituted, the difference would be greater: F1 had 726 Hz for males, 921 Hz for females, F2 was 1346 Hz for males and 1564 Hz for females, as the Polish /a/ is far lower and slightly more retracted than the second element of the /eə/ diphthong.

The /ʊə/ diphthong phoneme in native English can be pronounced as a diphthong or it can be monophthongized to /ɔː/. The monophthongization process is not a typical English vowel reduction as it takes place in stressed positions, and it does not result in a schwa, but a long mid back vowel /ɔː/. The problem is that the monophthongization process does not apply to all /ʊə/ phonemes (see Section 2.1). The /ʊə/ diphthong is hypothesized to be pronounced as /uɛ/, /ua/ or, to avoid a hiatus, with an inserted glide /uwɛ/ or /uwa/. The first element of the /ʊə/ diphthong is expected to have a quality of Polish /u/, so instead of being characterized by F1 of 426 Hz for males and 420 for females, and F2 of 1028 Hz for males and 1157 Hz for females, its expected F1 is 360 Hz for males 454 Hz for females and its expected F2 is 787 Hz for males 834 Hz for females. The comparison for the first element is not conclusive, because F1 for Polish /u/ is lower for males and higher for females than the F1 for the first element of the /ʊə/ diphthong. According to the articulatory descriptions, Polish /u/ is higher than English /ʊ/, so F1 of the /ʊə/ diphthong produced by Poles should actually be lower than if a native speaker produced it. The second element of the /ʊə/ diphthong is reported by Cruttenden (2001) to have F1 at 587 Hz for males and 485 Hz for females, whereas F2 at 1250 and 1258 Hz for males and females respectively. Polish /ɛ/ is at 537 Hz for males and 602 Hz for females, so the height seems to be similar, but the Polish vowel is fronter having a higher F2: 1862 Hz for males and 2084 Hz for females. Polish /a/ has F1 at 726 Hz for males and 921 Hz for females, and F2 at 1346 Hz for males and 1564 Hz for females, so it is much lower and slightly less retracted than the English /ə/.

4. Method

4.1. Materials

This study focuses on the quality of three British English centering diphthongs produced by Polish learners of English. The diphthongs were tested in real words only. Alveolar consonants were favored as an environment for the tested diphthongs because, as Lehiste and Peterson (1961) showed, the range of second formant fluctuation is relatively small. The chosen words were embedded in sentences. In the experimental design it was taken into account that segment durations are influenced by factors such as: speaker, intrinsic segmental properties, segmental context, prosodic context, and the global rate of speech (Klatt, 1976).

Intrinsic segmental properties were independent variables of interest. Segmental context was controlled by choosing the words in which diphthongs were surrounded by alveolar stops, or if impossible, fricatives, and if no other context was possible, by other classes of sounds. As for the prosodic context, an effort was made for the sentences to contain similar segmental material and similar number of syllables.

In order to control for rate of speech, initial practice sessions and randomization of test materials within the three blocks, each containing the set of 61 sentences, was used. Initial trial sessions were conducted to ensure that the talkers do not increase the rate of speech once they get familiar with the type of the task. The sentences were randomized within the blocks to counterbalance the order across the three blocks read by one subject and to counterbalance the order across subjects.

4.2. Subjects

There were nine male speakers. The age of the subjects ranged from 19 to 25, with the mean age of 22 years. All the subjects spoke English at an advanced level. They passed the Cambridge First Certificate Examination and at the time of the recording were preparing for the Cambridge Certificate in Advanced English. None of the subjects had ever received pronunciation training or had been to an English speaking country for more than a month. All the subjects received instruction in British English - their teachers were Poles aiming at speaking British English or occasionally for some period of time British native speakers and the coursebooks and tapes the subjects were taught from were also British English oriented.

In order to enable a differential study of diphthong patterns in triangulation three types of recordings were taken. The focus of the experiment was English diphthong production by Polish learners of English, so Polish learners of English were recorded reading English sentences. Since Polish vowel-plus-glide sequences were hypothesized to be the source model for Polish learners producing English diphthongs, recordings of Polish vowel-plus-glide sequences were made to allow for evaluating which processes typical for Polish vowel-plus-glide sequences Polish learners of English retained in English diphthong production. A native speaker of Polish read a set of Polish sentences prepared so as to examine Polish vowel-plus-glide sequences in contexts comparable to

the ones in which the English sentences were recorded. Since English diphthongs are the target models for English diphthong production by Polish learners of English, recordings of native English diphthongs were made to allow for estimation whether the Polish learners of English acquired English diphthong categories and associated processes. A native speaker of English was recorded reading the same English sentences that Polish learners of English read. For the clarity of presentation, the following abbreviations will be used.

- the native Polish learners of English will henceforth be abbreviated as PLE, and the corpus obtained from the Polish learners of English - PLEC
- the native speaker of English will henceforth be called NSE, and the corpus obtained from this speaker - NSEC
- the native speaker of Polish will henceforth be called NSP, and the corpus obtained from this speaker - NSPC

4.3. Procedure

The recording scenario involved diphthongs embedded in 61 sentences, each read three times by each subject. The subjects were instructed to read the sentences at a normal speed and with a falling intonation. The subjects controlled the tempo of recordings themselves and they were allowed to repeat a sentence when they wished to do so. The sentences were displayed on the computer screen in random order. The recordings were made with a 22050 Hz sampling frequency and a 16-bit resolution in a quiet office environment.

4.4. Annotation

The data were hand-annotated with Praat (Boersma and Weenink, 2007), using the SAMPA (Wells, 1997) phonetic alphabets for Polish and English, with an orthographic tier, and then with a tier containing segments of interest and their contexts. These were annotated with broad and narrow transcriptions and the canonical British English transcription was also noted for each word containing a diphthong of interest. The corpus is stored in the XML format, with TASX specifications. The re-usable format allows for further applications of the corpus. The corpus has been recorded, annotated and stored following EAGLES recommendations (Gibbon et al., 1997, 2000).

4.5. Measurements

The diphthong measurements which allowed for obtaining results needed to test the hypothesis of the paper were taken. In order to examine diphthong duration and duration relations within a diphthong the following measurements were made:
- the duration of the whole diphthong
- the duration of the steady state of the first target
- the duration of the steady state of the second target

In order to analyze the spectral qualities of a diphthong the following measurements were made:

- the frequencies of the first three formants at the first target steady state
- the frequencies of the first three formants at the second target steady state (more targets were analyzed if the PLEC realization contained more than two targets)
- the frequencies of the first three formants at the points set at 20% and 80% of the total diphthong duration

As in Stevens (1989, p. 10), the principal concern is with frequencies extending up to 3-4 kHz, i.e. F1, F2, and F3. It is because the auditory resolution at higher frequencies, expressed in terms of the width of the critical bands, is poor and listeners may be insensitive to the spacing of spectral prominences. Moreover, as a result of increased radiation losses, the acoustic losses at higher frequency ranges are greater. Individual formants cause less prominent spectral peaks and changes in the frequency of peaks do not significantly influence the overall spectrum shape. These measurements made it possible to obtain:

- information about diphthong duration
- information about the duration of diphthong phases
- information about the formant frequencies
- data for calculating the rates of frequency change associated with each formant movement
- vector length

4.6. Segmentation

In order to ensure control of prosodic variables and yield reliable duration and frequency measurements, segmentation was based on methodological principles of reliable and accurate acoustic speech segmentation (cf. Turk et al. (2006)). Segmenting the speech signal is an artificial task, because speech signal is inherently continuous and consists of overlapping gestures. Segmenting the L2 speech signal is even more challenging for the annotator than segmenting the speech signal in their L1, because there is a interplay of L1, L2 and interlanguage cues. As a result many new sound qualities, which are functions of L1, L2 and interlanguage, are created. The underlying assumption is that all acoustic landmarks could potentially be deciphered and that there is always a reason why they appear, it is the annotator's task to decipher as many cues as possible. The annotator, whose perception is always, to some degree at least, language-specific, has to ensure that the acoustic landmarks are consistently interpreted. Hence the need for precise segmentation rules.

Usually acoustic segment durations are defined in terms of oral consonantal constrictions intervals marked by abrupt spectral changes, and not the onset or offset of voicing (Turk et al., 2006). In the annotation of the present corpus, the voicing criterion was not used either, because comparable criteria were needed for voiced and voiceless consonants. As for consonantal constrictions, important as they were in vocalic boundary assignment, it was also ensured that strong coarticulatory effects or the aspirated or spirantized parts of the speech signal are not included within the diphthong boundary so

as not to distort the formant frequency measurements. Vocalic signal purity overrode the importance of consonantal constrictions. Traces of formant transitions, aspiration and burst noise which cues the identity of the surrounding consonants were not included within the vocalic boundaries, because the aim of the study was exclusively vocalic elements, and because there was a variety of consonantal surroundings. The variety of consonantal surroundings was limited where possible, but it was permitted so that diphthongs could be tested in real words. This methodology of disregarding VOT, bursts, constrictions and onset of voicing makes the segmentation employed here improper for reusing for the purpose of duration comparisons between voiced and voiceless obstruents.

Decisions on segmentation were primarily based on inspection of wideband spectrograms, combined with auditory perception, and with reference to the waveform in case of doubt. The formant trackers available in Praat were also consulted. Zoomed-out spectrogram displays were used to determine general boundary regions, and more zoomed-in waveform displays were used for marking exact boundary locations.

The choice of alveolar stops as the diphthong environment permitted relatively easy segmentation based on a decrease in amplitude and cessation of all but the lowest formant and harmonic energy (Turk et al., 2006). Probably because of L1 background of the subjects, there were no problems with variability of phonetic realizations of stops, there were no cases of glottalization before final stops and no cases of tapping.

It was challenging to segment diphthongs followed by /r/ sounds. Although, it was not the aim of the study to examine /r/ sounds, it turned out that they frequently occur as a part of a centering diphthong. Polish /r/ is a dental tap, whereas English /ɹ/ is a post-alveolar approximant. The quality produced by Polish learners of English is yet a third quality. Turk et al. (2006) also consider approximant segmentation notoriously difficult. Reliance on the midpoint of transitions from a preceding and to a following vowel is rather difficult to apply when vowels lack clear steady states. Additionally, it is not transparent to which articulatory events these transition midpoints belong, because they do not correspond to points of constriction onset and release, which make stops and fricatives easier to segment.

Generally, it was fairly clear where the diphthong boundaries after and before stops occur. There were also less clear cases, where it was however certain that the boundaries occur within a short window of uncertainty defined roughly as the duration of a single pitch period, i.e. 5-10 ms. Such cases were annotated throughout the dataset according to the chosen policy "when in doubt place the boundary in such a way as to ensure a relatively pure vocalic quality" (cf. Turk et al. (2006)). Many cases of diphthong plus /r/ or diphthong plus nasal sequences were judged to have boundary locations somewhere within a relatively wide window of uncertainty (i.e. more than 10 ms). It was then decided to include the /r/ into diphthong measurements, and longer durations of diphthongs in these contexts.

This strategy of including more segments than just the pure diphthong permitted relatively reliable measurements of duration for diphthongs and sounds so closely associated with them by Polish learners of English that it was impossible to tell them apart. Such a step has consequences for measurements of diphthong duration. Diphthong durations in open syllables and in syllables closed a voiceless consonant in centering diphthongs are longer than if only the diphthongal part had been measured.

Segmentation strategy resulting in influencing duration is permitted, as clipping effects are found in English vowels as well as in approximants. Thus it is assumed that including approximants in measurements will not interfere with finding out whether Polish learners of English have acquired the process of clipping or not.

The most important principle in segmentation was using the adopted criteria consistently, so that the results of segmentation could be compared. The target materials, i.e. diphthong surroundings were chosen carefully and carrier sentences were rahter similar with reference to prosodic criteria.

5. Production of centering diphthongs by Polish learners of English

This section reports, analyzes and interprets the results of an experiment on English centering diphthong production, including a number of specific phenomena, as for example breaking (i.e. /j/ or /w/ insertion), and /r/ insertion.

5.1. The /ɪə/ diphthong

In the realization of the /ɪə/ diphthong we can distinguish three basic types:
- consisting of four distinguishable qualities: /ijɛr/ or /ijər/ - 9 occurrences
- consisting of three distinguishable qualities: /jer/ or /jər/ - 7 occurrences
- consisting of two distinguishable qualities: /ie/ or /iə/ - 2 occurrences

If in the PLEC the /ɪə/ diphthong realization had four distinguishable qualities, the mean duration of the first target was 36 ms long (sd = 9 ms), of the second target 63 ms (sd = 17 ms), of the third one 54 ms (sd = 21 ms), and of the last one 53 ms (sd = 20 ms), thus the first target may be analyzed as an onglide, the second target being the longest one, and the remaining two targets being relatively long. If the realization had three distinguishable targets, the first target was 54 ms long (sd = 18 ms), the second one 77 ms (sd = 17 ms), the third one 54 ms (sd = 15 ms), with the total duration of the realization being approximately 20 ms longer than in the case of a four-target realization. A three-target realization seems to have one relatively long target being surrounded by two slightly shorter ones. A two-target realization of the /ɪə/ diphthong had the first target of 84 ms and the second target of 50 ms. Only the two-target realizations can be reasonably compared to the NSEC model, whose first target was 86 ms long and the second one was 83 ms long. No standard deviations are given for the two-target realizations, as both the NSPC and NSEC means were based on two measurements only.

Table 4. The /ɪə/ diphthong quality.

Type	Target no.	F1	F2	F3
4-target	1	328	2171	2777
4-target	2	339	2074	2582
4-target	3	370	1621	2858

Type	Target no.	F1	F2	F3
4-target	4	387	1509	2305
3-target	1	342	1920	2591
3-target	2	372	1740	2410
3-target	3	380	1660	2368
2-target	2	295	1705	2525
native	1	474	1814	2577
native	2	490	1668	2535

As the quality of the /ɪə/ diphthong presented in Table 4, in the four- and three-target realizations, F1 rose, whereas F2 and F3 decreased. In the two-target realization, F1 and F2 were relatively similar, and it was only F3 which was lowered, but not to the extent to which it was lowered in the case of the four- and three-target realizations ending with /r/-quality. The comparison of the two-target realization of the /ɪə/ diphthong by the PLEC speakers with the realization by the NSPC speaker revealed that the production by the PLEC speakers had a lower F1, but similar F2 and F3. The vector length for the two-target production by the PLEC speakers, which was 184 Hz was relatively similar to the NSPC value, which was 216 Hz. The vector lengths for four- and three-target realizations were 630 and 270 Hz respectively.

5.2. The /eə/ diphthong

The English /eə/ diphthong was realized by the PLEC speakers as an /er/ sequence, and it has to be noted that the /r/ component was of an auditory quality that resembled neither the Polish nor the English consonant. As far as timing relations are concerned, the PLEC speakers needed 135 ms (sd = 45 ms) to produce the first target and 111 ms (sd = 52 ms) to produce the second target. In the NSPC sequence /er/, the vowel took up 116 ms and the /r/ took up 25 ms. The first target durations were not significantly different in the NSPC and in PLEC, whereas the /r/ duration was significantly longer in the PLEC. The NSE happened to have a monophthongal realization of the /eə/ diphthong, which is consistent with the tendencies described by Cruttenden (2001) and Upton (2003). Therefore the results between a two-target realization by the PLEC speakers can hardly be compared with a one-target realization by the native speaker. Nevertheless, it should perhaps be mentioned that the vowel in the NSEC lasted 168 ms and its first three formants had the following values: 712 Hz, 1615 Hz and 2467 Hz.

The remaining discussion of the /eə/ diphthong will be based on the data from the PLEC and NSPC, with references to the NSEC monophthongal realizations when they will be feasible. The mean F1 value for the first target for learners was 585 Hz, whereas for the NSPC speaker it was 529 Hz, and the difference was not statistically significant. The values slightly resembled those of the first target formant for the /eɪ/ diphthong, which were 468 Hz for the PLEC speakers and 532 Hz for the NSP. In the case of /eə/, the F1 for the PLEC speakers was higher than the F1 for the NSP, which was a positive

result if we take into account that F1 should be raised by the PLEC speakers (although its value was still statistically significantly different from the monophthongal realization of F1 in NSEC, which was 712 Hz, if we are allowed to make such a comparison).

The value of F2 for the first target was 1579 Hz in the PLEC and 1581 Hz in the NSPC, and the difference was not statistically significant. The difference between the F2 in the PLEC and the F2 in NSEC (1615 Hz), was not statistically significant, either. F3 value is 2330 Hz in the PLEC, 2561 in the NSPC and 2467 Hz in the NSEC. The difference was significant between the F3 in the PLEC and in the NSPC, but not between the F3 in the PLEC and the F3 in NSEC.

The value of F1 for the second target was 439 Hz in PLEC and 383 Hz in NSPC, and the difference was not statistically significant. The values for F2 are also similar, i.e. 1489 in the PLEC and 1459 in the NSPC. F3 had statistically different values in the PLEC (2231 Hz) and in the NSPC (2491 Hz). As mentioned it the beginning of this section, the auditory quality of the /r/ produced by the PLEC speakers and the /r/ usually produced by Polish native speakers, who do not have speech impairments, are very different. The /r/ sound found in the PLEC was neither a tap nor a trill, it resembled a spirantized sound. The realization of the sound in such a way might be interpreted as a result of the motoric difficulty with producing a native-like /ɹ/, and lack of precision. It would, however, require further investigations whether it is F3 that is responsible for the difference, or some other factors, like higher formants, F0. Nevertheless, the /r/ was not found in the tested context in British English, and even if we tried to compare learners' realization to the American English /ɹ/, it would turn out that the American /ɹ/ had all the formants lower than the values for Polish learners presented here, F1 is around 320 Hz, F2 is around 1100 Hz, but especially F3 is much lower, i.e. around 1600 Hz (Yavaş, 2006, p.122).

In terms of acquisition, the PLEC speakers should learn not to produce the /r/ sound as a part of or after the English /eə/ diphthong, unless linking or intrusive /r/ is contextually justified. In terms of formants, both F1 and F2 of the first target should be raised.

5.3. The /ʊə/ diphthong

In the realization of the /ʊə/ diphthong we can distinguish four basic types:
- consisting of five distinguishable qualities: /juwar/ - 1 occurrence
- consisting of four distinguishable qualities: /juar/ - 5 occurrences
- consisting of three distinguishable qualities: /jua/ - 1 occurrence and /uar/ - 1 occurrence
- consisting of two distinguishable qualities: /ur/ - 10 occurrences

Longer realizations were typical for the word *Stuart* /stʊət/, and it has to be admitted that in this case the pronunciation might have been guided by loanword phonology, according to which the name *Stuart* is pronounced as /stjuart/ in Polish. The segment /j/ might not have been the aim of the analysis, but palatalization resulting in /j/ was clearly

audible both in the NSEC (to a lesser extent) and in the PLEC (to a greater extent). Since palatalization is the effect of the influence of the /ʊ/ segment on the preceding /t/ segment, it is not clear with which segment it should be associated. In order not to miss important features of the diphthong, the /j/ was also taken into account, and the palatalized cases with the word *Stuart* are analyzed separately from the word *tour*. Shorter realizations are typical for the word *tour*. The realization in the NSEC was /tʊər/ because of the linking /r/ inserted to link the word with the following word *around*. In light of the fact that the linking /r/ applies in the word in such a context in native English, it should not be claimed in the case of the word *tour* realization in this experiment that the PLEC speakers should produce a schwa instead of the /r/ sound, but the emphasis could be on the diphthongal nature of the vocalic element.

Table 5. Duration (ms) of the targets in the 5-, 4-, 3- (for the word *Stuart*) and 2-target (for the word *tour*) realizations of the /ʊə/ diphthong in the PLEC, and in the realizations of words *tour* and *Stuart* in the NSEC and the Polish word *tur* in the NSPC.

Target	5-target	4-target	3-target	NSEC *tour*	NSEC *Stuart*	2-target	NSPC *tur*
1	48	50	96	81	55	115	138
2	77	70	99	72	78	96	113
3	45	85	86	72	42	x	x
4	147	105	x	x	x	x	x
5	55	x	x	x	x	x	x

A conclusion that can be drawn from Table 5 is that the realization of the English diphthong in the PLEC did not differ significantly from the realization of the word *tur* in the NSPC in terms of the first target duration (p = -2.23 for df = 9), and in terms of the relation between the duration of the first and second targets, which was 1.28 in the PLEC and 1.22 in the NSPC (p = 0.35 for df = 9). The /r/ segment was shorter when produced by the PLEC speakers than in the NSPC (p = -2.42 at 9 df).

Table 6. Quality of the targets in the word *tur* in the NSPC, and the realizations of *tour* and *Stuart* in the PLEC and NSEC.

Corpus	Word	Sound	F1	F2	F3
NSPC	*tur*	/u/	430	763	2236
NSPC	*tur*	/r/	412	1229	2107
PLEC	*tour*	/u/	393	1117	2364
PLEC	*tour*	/r/	445	1191	1969
NSEC	*tour*	/ʊ/	545	1384	2444
NSEC	*tour*	/ə/	557	1319	2120
NSEC	*tour*	/r/	330	1240	1379
PLEC	*Stuart*	/j/	328	1817	2457
PLEC	*Stuart*	/u/	405	1126	2267

Corpus	Word	Sound	F1	F2	F3
PLEC	*Stuart*	/a/	565	1196	2146
PLEC	*Stuart*	/r/	529	1446	2272
NSEC	*Stuart*	/j/	449	1908	2354
NSEC	*Stuart*	/ʊ/	503	1443	2217
NSEC	*Stuart*	/ə/	488	1397	2423

Table 7. Summary of t-tests for the quality of the /ʊə/ diphthong. Significant results (p=5) are denoted by boldface.

Test	Word(s)	Target	Formant	t value
PLEC vs. NSPC	*tour, tur*	/ʊ/	1	**-2.87**
PLEC vs. NSEC	*tour*	/ʊ/	1	**-12**
PLEC vs. NSPC	*tour, tur*	/u/	2	**7.48**
PLEC vs. NSEC	*tour*	/ʊ/	2	**-5.65**
PLEC vs. NSPC	tour, tur	/u/	3	1
PLEC vs. NSEC	tour	/ʊ/	3	-0.62
PLEC vs. NSPC	tour, tur	r	1	1.07
PLEC vs. NSEC	tour	/r/	1	**3.65**
PLEC vs. NSPC	tour, tur	/r/	2	-1.07
PLEC vs. NSEC	tour	/r/	2	-1.38
PLEC vs. NSPC	tour, tur	/r/	3	-1.41
PLEC vs. NSEC	tour	/r/	3	**6.06**

The discussion of the quality of the /ʊə/ diphthong is based on the ten two-target realizations of the diphthong in the PLEC, i.e. /ur/, which will be compared to the word *tur* in the NSPC, pronounced as /tur/, and the word *tour* in the NSEC, realized with the linking /r/ as /tʊər/. The discussion is based on the results shown in Tables 6 and 7.

In comparison with realizations in the NSPC and NSEC, in the PLEC the F1 for the first target of the diphthong had the lowest value, which was significantly different from the value in the NSPC and the highest value in the NSEC. This result means that the PLEC speakers did not acquired the right height for this vocalic element. It would perhaps be worthwhile to examine in the future how this F1 value relates to the F1 values of native speakers of English and Polish learners of English for the /uː/ monophthong, because of a possibility that it might be difficult for Polish learners of English to distinguish between /uː/ and /ʊ/. This hypothetical explanation seems to be justified in light of the results of perception tests showing that a group of Polish learners of English similar to the one examined in the present study did not perceive the difference between the two English high back vowel (Bogacka [Balas], 2004). It is thus feasible that if Polish learners of English are not able to perceive the difference between /uː/ and /ʊ/ and categorize the two vowels into one category, the less likely they are to

distinguish between similar /ʊ/ monophthong and the first part of the /ʊə/ diphthong and use the /u/ category for all three English vocalic qualities.

Results for F2 of the first target seem to indicate proper direction in the acquisition. The value in the NSPC is the lowest one, F2 was higher in the PLEC and the highest in the NSEC. In terms of statistical significance of the differences between F2 in the PLEC and the values in the NSPC and NSEC, they were statistically significant, but the difference between F2 in the PLEC and NSEC was smaller than in the case of F2 in the PLEC and NSPC. Thus it can be claimed that acquisition is advanced.

The results for F3 of the first target were proportionately similar to the results for F2, but the differences between the PLEC and NSPC or NSEC were not statistically significant. This is considered to be a desired outcome, since lip rounding, which is related to F3, is an important feature of high back vowels.

The second target of the /ʊə/ diphthong in the NSEC cannot be reasonably compared to any of the segments produced by the PLEC speakers. The PLEC speakers seem to be producing a kind of /u/ followed by /r/, when English /ʊə/, followed by the linking /r/ depending on context, should be produced. We can therefore compare the /r/ qualities. The first formant has the highest value for the realization by the PLEC speakers and the lowest value for the realization in the NSEC. The difference between the value in the PLEC and NSPC was not statistically significant, whereas the difference between the value in the PLEC and NSEC was statistically significant. It seems that, as in the case of the first target, the acquisition of F1 has not taken place. The differences between F2 values as produced by the PLEC speakers and NSE and NSP were not statistically significant, so there was no need for acquisition. F3 had the lowest value in the NSEC, the value in the PLEC was significantly higher, and the value in the NSPC was not significantly higher than the learners' F3. The results for F3 lead to the conclusion that F3, reflecting a very characteristic feature of the English /r/ sound, is very difficult for Polish learners to acquire.

6. Conclusions

No definite statements can be made about the acquisition of vocalic height, which seems to be very difficult. The F1 values for the first target of /ɪə/ ranged from 305 Hz to 342 Hz, whereas the NSEC value was 474 Hz. Although no statistical tests were carried out on such small samples, it was clear that the first part of the /ɪə/ diphthong in the NSEC had a higher value. In the /eə/ diphthong, F1 of the first target produced by the PLEC speakers was also not statistically different from the F1 value in the NSPC, but it was statistically different from a lower F1 value in the NSEC. Also in the case of the diphthong /ʊə/, the F1 for the first target was statistically significantly lower in the PLEC than in the NSEC or NSPC. The second targets of centering diphthongs produced in the PLEC, are distorted by breaking processes and /r/ insertion.

In terms of the direction of F1 change, which has not been acquired, there are F1 increases as in the case of the first targets of rising diphthongs, the first target of /ɪə/ and /ʊə/, and F1 decreases as in the case of the first target of /eə/.

The results for F2, related to the tongue advancement, were even less definite, as sometimes the direction of change is opposite to the expected one.

F2 for the first target of the /ɪə/ diphthong seemed to be similar to the NSEC model. F2 of the first target of the /eə/ diphthong was not statistically significantly different from the NSEC model. As for the first target of the /ʊə/ diphthong, there seemed to be acquisition in progress, as the F2 in PLEC was between the Polish and English values. It was statistically different from both NSPC and NSEC models, but closer to the NSEC model. The value in the NSPC was the lowest, F2 in the PLEC was closer to the F2 in the NSEC, which had the highest value.

As for the direction of change, the acquisition of F2 took place when F2 had to be raised in the first target of /eə/, though the raising did not need to be substantial, as the PLEC values did not significantly differ from the NSPC values. The hypothesis was that approximating the tongue height should be favored over approximating the tongue backness in the case of non-low front vowel region, as in this region F2, F3 and F4 are rather insensitive to anterior-posterior displacement of the tongue body (Stevens, 1989). It was thus expected that the first targets of /ɪə/ and /eə/, there should be a preference for approximating F1 values to the English F1 values, over approximating F2 values. In /ɪə/ it cannot be stated, because of so many various realizations of the diphthong. In /eə/, there was a tendency for raising F1 towards the NSEC value, but the difference between the two values was still statistically significant. F2 values were very similar for the three groups of speakers, and the value in the PLEC was not statistically significantly different from neither of the model values.

Unexpectedly, the PLEC F3 values in both the first and second diphthong targets were, in the majority of cases, statistically significantly different from the F3 values in the NSPC, whereas they were not statistically significantly different from the F3 values in the NSEC. This result was not expected because F3 values were not considered to play a decisive role in vowel perception. Thus it was not expected that learners would make an effort to acquire a feature which is not vital in communication. Perhaps, however, in order to produce more English-like diphthongs, it is easier to begin changing features which do not distort the vowel space system of L1 to a large extent.

References

Boersma, P. and D. Weenink. 1992-2008. *Praat: doing phonetics by computer.* Computer program. Retrieved from http://www.praat.org/.

Bogacka [Balas], A. 2004. On the perception of English high vowels by Polish learners of English. In E. Daskalaki, N. Katsos, M. Mavrogiorgos, and M. Reeve (eds.), *CamLing 2004: Proceedings of the University of Cambridge Second Postgraduate Conference in Language Research.* Cambridge: Cambridge Institute of Language Research. 43 - 50.

Bogacka [Balas], A., G. Schwartz, M. Połczyńska-Fiszer, P. Zydorowicz, and P. Orzechowska 2005. The production and perception of schwa in second language acquisition: The case of Polish learners of English. In K. Dziubalska-Kołaczyk (ed.) IFAtuation: A Life in IFA. A Festschrift for Professor Fisiak on the Occasion of his 70th Birthday, pp. 71–84. Poznań: Wydawnictwo Naukowe UAM.

Cruttenden, A. 2001. Gimson's Pronunciation of English (6 ed.). London: Arnold.

Donegan, P. 1985. *On the Natural Phonology of Vowels*. New York: Garland Publishing, Inc.

Donegan, P. 1993. On the phonetic basis of phonological change. In C. Jones (ed.) *Historical Linguistics: Problems and Perspectives*. London: Longman. 98-130.

Donegan, P. 2001. Constraints and processes in phonological perception. In K. Dziubalska-Kołaczyk (ed.) *Constraints and Preferences*. Berlin: Mouton de Gruyter. 42-68.

Donegan, P. and D. Stampe. 1979. The Study of Natural Phonology. In D. A. Dinnsen (ed.) *Current Approaches to phonological Theory*. Bloomington: Indiana University Press. 126-173.

Dressler, W. U. 1984. Explaining Natural Phonology. *Phonology Yearbook* 1. 29-50.

Dressler, W. U. 1985. *Morphonology: the Dynamics of Derivation*. Ann Arbor: Karoma Publishers.

Dressler, W. U. 1996. Principles of naturalness in phonology and across components. In B. Hurch and R. Rhodes (eds.) *Natural Phonology: The state of the art,*. Berlin: Mouton de Gruyter. 41–52.

Dressler, W. U. 1999. On a semiotic theory of preferences in language. In M. Jaley and M. Shapiro (eds.) *The Peirce Seminar Papers. Essays in Semiotic Analysis. Proceedings of the International Colloquium on Language and Peircean Sign Theory*, 1997, vol. 4. New York: Berghahn Books. 389-415.

Dziubalska-Kołaczyk, K. 1990. *A Theory of Second Language Acquisition within the Framework of Natural Phonology*. Poznań: Adam Mickiewicz University Press.

Dziubalska-Kołaczyk, K. 1995. *Phonology Without the Syllable. A Study in the Natural Framework*. Poznań: Motivex.

Dziubalska-Kołaczyk, K. 2001. Phonotactic constraints are preferences. In K. Dziubalska-Kołaczyk (ed.) *Constraints and Preferences*. Berlin: Mouton de Gruyter. 69-100.

Dziubalska-Kołaczyk, K. 2002a. *Beats-and-Binding Phonology*. Frankfurt am Main: Peter Lang.

Dziubalska-Kołaczyk, K. 2002b Challenges for Natural Linguistics in the twenty first century: A personal view. In K. Dziubalska-Kołaczyk and J. Weckwerth (eds.) *Future Challenges for Natural Linguistics*. Munich: Lincom. 103-128.

Gibbon, D., I. Mertins, and R. Moore. 2000. *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. Dordrecht: Kluwer Academic Publishers.

Gibbon, D., R. Moore, and R. Winski (eds.). 1997. *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: Mouton de Gruyter.

Klatt, D. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59, 1208–1221.

Lehiste, I. and G. E. Peterson. 1961. Transitions, glides and diphthongs. *Journal of the Acoustical Society of America* 33, 268–277.

Liljencrants, J. and B. Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48, 839–862.

MacCarthy, P. 1978. The Teaching of Pronunciation. Cambridge: Cambridge University Press.

Majewski, W. and H. Hollien. 1967. Formant frequency regions of Polish vowels. Journal of the Acoustical Society of America 42, 1031–1037.

Polivanov, E. 1931. La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4, 79–96.

Ritt, N. 2001. Are optimality theoretical "constrains" the same as natural linguistic "preferences"? In K. Dziubalska-Kołaczyk (ed.) *Constraints and Preferences*. Berlin: Mouton de Gruyter. 291-310.

Roach, P., J. Hartman, and J. Setter (eds.). 2006. Cambridge English Pronouncing Dictionary. Cambridge: Cambridge University Press.

Selinker, L. 1972. Interlanguage. *International Review of Applied Linguistics* 10, 209–231.

Sobkowiak, W. 2004. *English Phonetics for Poles* (3 ed.). Poznań: Wydawnictwo Poznańskie.

Stampe, D. 1969. The acquisition of phonetic representation. *CLS* 5: 443-453.

Stampe, D. 1979. *A Dissertation on Natural Phonology*. New York: Garland Publishing, Inc.

Stampe, D. 1984. On phonological representations. *Phonologica*: 287–300.

Stevens, K., J. Keyser, and H. Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In J. Perkell and D. Klatt (eds.) *Invariance and variability in speech processes*. Hillsdale, NJ: Erlbaum. 426–449

Stevens, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17, 3–46.

Trubetzkoy, N. S. 1939/1969. *Principles of Phonology*. Berkeley: University of California Press.

Turk, A., S. Nakai, and M. Sugahara. 2006. Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schließer (eds.) *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter. 1–28.

Upton, C., W. K. Jr, and R. Konopka (eds.). 2003. The Oxford Dictionary of Pronunciation for Current English. Oxford: Oxford University Press.

Wells, J. C. 1997. Sampa computer readable phonetic alphabet. In D. Gib-bon, R. Moore, and R. Winski (eds.) Handbook for Standards and Resources for Spoken Language Systems. Berlin and New York: Mouton de Gruyter. Part IV, Section B.