DYNAMIC TARGETS IN THE ACQUISITION OF L2 ENGLISH VOWELS[*]

GEOFFREY SCHWARTZ
Uniwersytet Adama Mickiewicza w Poznaniu
geoff@wa.amu.edu.pl

GRZEGORZ APERLIŃSKI
Uniwersytet Adama Mickiewicza w Poznaniu
gaperlinski@wa.amu.edu.pl

KAMIL KAŹMIERSKI
Uniwersytet Adama Mickiewicza w Poznaniu
kamil.kazmierski@wa.amu.edu.pl

JAROSŁAW WECKWERTH
Uniwersytet Adama Mickiewicza w Poznaniu
wjarek@wa.amu.edu.pl

Abstract
This paper presents acoustic data on the dynamic properties of the FLEECE and TRAP vowels in the speech of two groups of Polish users of English. Results reveal that the more proficient group users, made up of teachers and professors with professional-level proficiency in English, produce more dramatic patterns of formant movement, reminiscent of native productions, than first year students. It is argued that vowel inherent spectral change (VISC) is an inherent aspect of English phonology, originated in interactions between vowels and neighboring consonants, and later generalized to the vowel system as a whole. By contrast, Polish is a language with a minimal role of VISC. Consequently, successful acquisition of L2 English vowels involves not only the mastery of vowels in F1-F2 space, but also formant trajectories over time.

Keywords: Phonetics, Phonology, Vowel Inherent Spectral Change, L2 English

1. Introduction - English vowels and L2 learners

The English vowel system presents second language (L2) learners with a number of challenges that complicate the process of target language acquisition. Regardless of their first language (L1) background, those seeking to achieve a high level of performance in their English pronunciation must deal with the fact that the L2 they are learning is characterized by a great deal of variability in the vowel system across different native accents. Differences may be found in the choice of vowel phoneme that may appear in a given class of words (e.g. words such as *bath* in British vs. US English: /bɑːθ/~/bæθ/). Alternatively, a vowel that is arguably the same phoneme across accents may be realized differently (e.g. *pot* [pʰɒt] vs. [pʰɑt]), potentially leading to learner uncertainty about the quality of a given vowel. This problem is exacerbated both by pervasive dialectal variation and sound changes in progress. For example, the TRAP vowel shows a wide range of realizations, including variable degrees of raising with schwa-like offglides in American English, and more open and retracted productions by younger generations of British speakers (Hawkins and Midgley 2005). [1] A more fundamental issue is the fact that the English vowel system is quite dense from a typological standpoint, and includes a number of uncommon vowels (e.g. /æ ɒ ʌ/), and difficult contrasting pairs (e.g. /uː/ vs. /ʊ/). Thus, learners from most L1 backgrounds will have to learn to produce new vowel sounds, and to perceive acoustic distinctions that are absent from their first language.

In this connection, there is a large experimental literature devoted to L2 learners' acquisition of English vowels. These studies cover a range of L1s, including those with large vowel systems such as German (e.g. Bohn and Flege 1997), and L1s with sparse vowel systems such as Spanish (Escudero and Boersma 2004) and Polish (e.g. Rojczyk 2011). This literature has been instrumental in the formulation and refinement of the most influential current models of L2 speech acquisition: the Speech Learning Model (SLM; Flege 1995) and the Perceptual Assimilation Model (PAM; Best 1995, Best and Tyler 2007). Research inspired by these theories has examined both the perception and production of many intricacies of the English vowel system, including subtle differences in vowel quality, as well as duration. Some important findings have suggested that in L2 perception learners become perceptually desensitized by acoustic cues from their L1 (Bohn 1995), a concept closely related to equivalence classification (Flege 1987). In these cases, it has been found that learners make use of other cues that are absent from their L1 system in the perception (and production) of new L2 categories. One notable example is the

---

[1] General confusion about the quality of this vowel may contribute to difficulties for learners. Interestingly, borrowings of the TRAP vowel into Polish are sometimes with /ɛ/ (*menedżer* 'manager') or /a/ (*faks* 'fax'), even for the same source word (*flesz* 'flash lamp' vs. *Flash* /flaʃ/ 'graphics software platform').

greater use of duration over formant cues by speakers without L1 duration contrasts in the realization of difficult English vowel contrasts (Bohn 1995; Escudero and Boersma 2004; Rojczyk 2011).

To the casual observer it might seem as if the topic of L2 English vowel acquisition has been exhaustively researched, and that there is little room for significant additional experimental investigation. However, there is an important current of experimental work into the phonetics of vowel quality that has remained somewhat outside of the mainstream for L1 English, and has gone almost completely unexplored in L2 speech research. This area investigates Vowel Inherent Spectral Change (VISC; e.g. Morrison and Assmann 2013), changes in vowel quality over the course of vowel duration, and its role in many areas, including production (Morrison 2013; Williams and Escudero 2014), perception (Hillenbrand 2013), and dialectal and diachronic variation (Fox and Jacewicz 2009). VISC research has found that for an adequate description of the English vowel system, traditional two-dimensional vowel charts are not sufficient, and that a deeper understanding is gained when dynamic patterns of formant movement are documented. Given the diachronic instability of the English vowel system (Hawkins and Midgley 2005; Labov et al. 2006, Kaźmierski 2015), this is to be expected – when vowel quality is dynamic, shifting and instability are a natural result.

At the same time, it appears as though patterns of VISC are not universal, but differ systematically across languages. This fact opens up a rich testing ground for additional research into L2 speech acquisition and cross-language comparison. Languages can be compared with regard to the degree (and/or directions) of spectral change found in their vowel systems, and L2 learners may be studied in terms of the extent to which they master or fail to master patterns of formant dynamics in the second language. For example, Polish is a language with a small vowel system comprised of vowels that are relatively stable in quality. Thus, it may be expected that Polish learners' acquisition of English vowels may be described in terms of patterns of VISC as a supplement to traditional two-dimensional vowel charts. For example, Schwartz (2010) suggested that Polish learners' [ɪ]-like productions of English /iː/ may be attributable to a lack of perceptual attunement to diphthongal realization of the target vowel. That is, learners may base their /iː/ percept on the early portion of the vowel in which the target has not yet been reached, and the formants indeed resemble typical values for /ɪ/.

This paper will present an acoustic comparison of vowel quality, with particular focus on VISC, in the speech of two groups of Polish speakers of English. More proficient speakers are hypothesized to exhibit more robust patterns of VISC that resemble native norms (for Southern British English), while less proficient learners are expected to produce L2 English vowels that are characterized by a smaller degree of formant movement. However, before describing the acoustic study, it is necessary to provide a more detailed

introduction to Vowel Inherent Spectral Change, with a focus on its possible origins, as well as links to other areas of English phonology.

2. The phonological origins of VISC

In this section, we will explore the hypothesis that the degree of vowel inherent spectral change in a given language is in essence a phonological consideration, not simply a phonetic detail. In the case of English, we may show links between vowel reduction and rhythm, spectral dynamics, and consonant place perception. In these areas Polish and English appear to show polar oppositions that make for robust cross-linguistic comparison. The discussion here briefly summarizes arguments outlined in Schwartz (2015). For a more thorough discussion of these issues, see Schwartz (2016).

It may be suggested that English-Polish oppositions in VISC reflect differences in rhythmic organization, at least in terms of its manifestation in vowel systems. Donegan and Stampe (1983) offer a perspective on rhythmic organization in language. They note that stress-timed languages typically are characterized by what they call 'dynamic vocalization', a tendency for diphthongization of monophthongs, as well as the presence of diphthongs. Kaźmierski (2014) found some quantitative support for the link between stress-timing and the presence of diphthongs in a language, as well as the diachronic instability of the vocalic system. The former emerged in a typological survey of 33 languages from 8 language families, in which diphthongs were present almost exclusively in stress-timed languages. The latter follows from the comparison of the prevalence of vocalic changes in three Germanic languages (all stress-timed, numerous vocalic changes), to their prevalence in Romance, Slavic and Finno-Ugric (mostly syllable-timed, significantly fewer vocalic changes).

Other research has suggested that 'dynamic vocalization' may be crucial for vowel perception in stress-timed languages. L1 English listeners appear to have an easier time identifying vowels spoken in various consonantal contexts than they do in isolation (e.g. Strange et al 1983, Jenkins and Strange 1999). Such findings have led to a theory of 'dynamic specification' in vowel perception (see Strange 1989), by which VISC, rather than static acoustic targets, constitute the primary cues for vowel identification. Considering the 'dynamic specification' theory in light of Donegan and Stampe's rhythmic typology, the degree of 'dynamic vocalization' should be a phonological product of a language's prosodic organization. Thus, we should expect dynamic specification, and VISC, to be less robust in syllable-timed languages than in stress-timed languages.

Dynamic specification research has emphasized the contextual effects of neighboring consonants on vowel formants (Jenkins and Strange 1999). If vowel dynamics are related to rhythm, and dynamic specification stems from consonant-vowel interaction, we should therefore expect more robust CV interactions in English, a stress-timed language, than in Polish, a syllable-timed

language.[2] This claim is also compatible with the findings of Święciński (2004), who compared articulatory settings in Polish and English for the purposes of teaching English pronunciation to Poles. Święciński (2004) describes a difference in the tension of the supralaryngeal articulators (the tongue and the lips) by which the English setting is lax (cf. Laver 1994), while Polish is characterized by a 'slightly tense' supralaryngeal setting. It may be hypothesized that the more relaxed articulatory setting in English contributes to slower transitions between vowels and neighboring consonants, leading to more robust formant movement in vowel realization. Another prediction that falls out from this discussion is that Polish and English should differ in the relative weight of CV formant transitions and aperiodic noise in the perception of consonant place of articulation. Since Polish vowels are purer, less of the vowel's duration is available to listeners as a cue to consonant place. Consequently, Polish listeners should weight noise cues more heavily than English listeners

These predictions were tested in a series of cross-linguistic and acquisition experiments, including Schwartz et al. (2014), Schwartz and Aperliński (2014), and Aperliński and Schwartz (2015). The first of these studies examined the relationship between stop release suppression and VC formant transitions in the speech of Polish learners of English. It may be claimed that the tendency for unreleased coda stops in English stems from the perceptual recoverability of stop place from the formant patterns on the preceding vowel. That is, since VC transitions are perceptually robust in English, speakers may spare the effort of producing release bursts without sacrificing 'sufficient discriminability' (Lindblom 1990). By contrast, in Polish coda stops are always released (Dukiewicz and Sawicka 1995), except in homorganic clusters, in which release may be suppressed. Schwartz et al. (2014) found that more proficient Polish speakers of English produce VC transitions that occupy a greater percentage of vowel duration than less proficient speakers, and are more likely to produce unreleased stops. The other studies (Schwartz and Aperliński 2014; Aperliński and Schwartz 2015) compared stop place perception in CV contexts by English and Polish listeners. They found that the relative perceptual weight of CV formant transitions for perception of consonant place of articulation was greater in English than in Polish. Polish listeners, by contrast, attended to the spectrum of noise bursts to a larger degree.

In most contemporary theories of phonology, considerations such as VISC and the relative weight of transition vs. burst cues to consonant place of articulation would be described as gradient phonetic details that fall outside the

---

[2]   On the basis of its rich inventory of consonant clusters, some authors (e.g. Ramus et. al 1999) have suggested that Polish is not a true syllable-timed language. See White and Mattys (2007) and Schwartz (2010) for arguments that vowel reduction metrics better capture perceived rhythmic categories than phonotactic metrics. See also Wagner (2007) and Malisz (2013) for evidence that, at least for spontaneous speech, Polish shows syllable-timing in coupled oscillator models.

realm of phonology proper. However, the connection with rhythmic classes mentioned above suggests that such cross-language differences are the product of categorical parameter settings. These parameters are built into the Onset Prominence phonological framework (OP; Schwartz 2013, 2016), in which the initial vocalic portion in CV sequences, represented as the Vocalic Onset (VO) node of structure (Schwartz 2013), is ambiguous with regard to the traditional consonant-vowel distinction. Phonetically, this portion of the signal is, strictly speaking, part of the vowel. Yet since it typically contains acoustic information about the identity of the preceding consonant, it may be built into consonantal representation. This ambiguity creates a parameter setting by which languages may differ in a systematic way. When the VO node is included in obstruent representations, we should expect more robust CV interaction, and dynamic formant patterns in vowel quality, as has been observed for English. In other words, dynamic vocalization is greater when the initial portion of vowels is contained in consonant representations. By contrast, in Polish, the VO node is posited to be part of vowel representations (see Schwartz 2013), with the effect that targets are reached earlier in the vowel and vowel quality is less diphthongal.

The OP representational parameters for Polish and English are illustrated in Figure 1, in which we see two different structural parses of a CV sequence. In both parses, the Closure (C) and Noise (N) nodes of the representational hierarchy are contained in the representation of the consonants, where they house place features. The difference between the two systems is based on the status of the VO. On the left, we see the VO node of structure contained in the lower-level vowel structure, dominating the Vocalic Target (VT). In this configuration, phonetic effects of the consonant (C-place) on the vowel quality are predicted to be minimal, since the phonological 'boundary' between consonant and vowel aligns with the acoustic boundary. In the configuration on the right, the consonant representation includes the VO, encoding CV formant transitions on the vowel in the representation of the consonant. As a consequence, vowel quality is more susceptible to the effects of consonant-induced co-articulation.
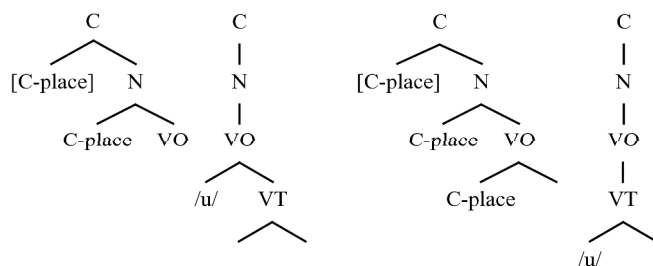
Figure 1. Onset Prominence representational parameters for stop-vowel sequences
(after Schwartz 2016)

As an illustration of cross-linguistic influence involving these representations, consider the English word *soon* as spoken by a native speaker of American English and by a speaker with a strong Polish accent. In the native realization, VO is contained in the representation of the consonant as we see on the right. In this configuration, the consonant impinges on the structure of the vowel, so the anterior coronal fricative /s/ in *soon* has a significant fronting effect on initial portion of the vowel, which is typically realized as [ʉ] at vowel onset. Only later in the vowel is the back tongue position reached. The fronting effect is typically evident in significant movement of the second formant (F2), which starts high in the spectrum and then gets lower. When *soon* is spoken with a heavy Polish accent, the vowel is almost instantaneously realized as a back vowel. The lack of fronting is predicted by the structures on the left, in which the consonant does not impinge upon the structure of the vowel. Thus, in the Polish realization, the anterior place of articulation of the consonant has a much smaller impact on the realization of the vowel, and the F2 frequency remains more or less constant over the course of the vowel.

The insight of OP representations lies in the fact that there is a built-in layer of prosodic structure that reflects the transition between consonant and vowel in CV sequences. The segmental affiliation of this structural node, either consonantal or vocalic, makes predictions about vowel dynamics. Traditional phonological representations are not capable of predicting the greater fronting effect of coronal consonants on vowels in English than in Polish. Thus, while recent works on English dialectology may describe the vowel in *soon* as fronted, the underlying origin of the fronting process in consonant-vowel interaction is left unexplained in most models.[3]

3. Acoustic study

This section will present a pilot acoustic study comparing two groups of Polish speakers of English in terms of spectral dynamics of L2 vowel production. The paper is intended to build on data presented in Schwartz (2015), employing additional automated measures of VISC that allow for a greater amount of data, greater reliability, and greater detail in acoustic description. As a preliminary report from a larger project on the production and perception of VISC by Polish learners of English, in this paper we only present results from two vowels.

---

[3] In many accents of English, the fronting of this vowel has spread beyond coronal contexts (Docherty 2010). However, data from other dialects, such as California English in which only coronals exhibit fronting effects (Ladefoged 1999), suggests that the fronting process must have originated in the coronal context.

3.1. Participants and materials

Twenty Polish speakers of English took part in the experiment. Participants were divided into two groups. The first group was comprised of students in their first year of studies at the Faculty of English at Adam Mickiewicz University in Poznań (Students). The second group (Teachers) was made up of Staff members at the Faculty of English, including both PhD students and professors. Thirteen of the twenty participants were female and seven were male. Both groups of participants had been instructed using a British pronunciation model.

Acoustic data were extracted from recordings of an English word list (see Data Collection and Analysis). The preliminary analysis presented here includes four repetitions of eight words containing two vowels, in four different consonantal contexts, summarized in Table 1.

Table 1. Summary of experimental items with consonantal context

Word	Vowel (keyword)	Onset	Coda
bead	/iː/ (FLEECE)	Unaspirated labial	Lenis coronal
beat	/iː/ (FLEECE)	Unaspirated labial	Fortis coronal
deed	/iː/ (FLEECE)	Unaspirated coronal	Lenis coronal
neat	/iː/ (FLEECE)	Unaspirated coronal	Fortis coronal
bad	/æ/ (TRAP)	Unaspirated labial	Lenis coronal
bat	/æ/ (TRAP)	Unaspirated labial	Fortis coronal
dad	/æ/ (TRAP)	Unaspirated coronal	Lenis coronal
stat	/æ/ (TRAP)	Unaspirated coronal	Fortis coronal

The choice of /iː/ and /æ/ for the preliminary analysis requires additional comment. The vowel /iː/ was chosen because in many dialects of English it has been described as diphthongal in quality – notably in the Southern British standard serving as the point of references for the speakers analyzed in this study. In other words, we may expect it to be characterized by a significantly greater degree of Vowel Inherent Spectral Change than is observed in the putatively 'similar' vowel /i/ in Polish. In the case of /æ/, the choice was made for two primary reasons. First and foremost, the vowel is notoriously difficult for Polish learners, who have been known to confuse it with /e/ and /ʌ/ in English, and to substitute it with Polish /ɛ/ or /a/ production (Weckwerth 2011; Gonet et al. 2010). Secondly, as an open vowel, /æ/ is expected to show more robust co-articulatory effects of neighboring consonants, particularly in the first formant, than other vowels. That is, since most consonants are not articulated with a low tongue position, the transitions to and from /æ/ will entail a significant amount of F1 movement. Since it is hypothesized that the origins of VISC may lie in consonant-vowel co-articulation, looking at the TRAP vowel provides us with a good vantage point from which to describe the relationship between VISC and neighboring consonants.

3.2. Data Collection and Analysis

Recordings were made in a laboratory setting in a sound-treated chamber at the Faculty of English, Adam Mickiewicz University in Poznań. An MXL-700 small diaphragm condenser microphone, placed about 30 cm from the speaker's lips, was used to capture the sound. It was plugged into an Edirol Duo Capture USB interface connected to a laptop computer. The participants read a wordlist of 64 English words representing 8 English vocalic phonemes traditionally described as "monophthongs" in 8 different consonantal contexts (only two vowels and four contexts are analyzed here. The list was divided into four blocks, and within each block the order of presentation was randomized for each target vowel. The stimulus items were presented, one at a time, on a computer screen, and elicited using the from presentation slides shown to the participants using the Speech Recorder program (Draxler and Jänsch 2015), which saves the recordings induced from each slide as a separate sound file; in our case, the files were PCM wave files at a sampling frequency of 44.1 kHz and bit depth of 16 bits. The Recordings were annotated by hand into Praat text grids by the first author of the study, according to standard criteria for measuring vowel duration (e.g. Lass, 1996).

A Praat script was used to automatically extract a number of acoustic measures relating to both static vowel quality and VISC. The script creates a Praat Sound object for each analyzed vowel on the basis of text grid boundaries. Calculations are then run on the Sound objects rather than in the viewer window to ensure reliable values across all measurement points.[4] The script also allows the experimenter to choose the number of intervals into which to divide a given vowel, and returns mean F1 and F2 values for each interval, normalized according to Syrdal and Gopal's (1986) Bark transformation (F1-f0 for vowel height, F3-F2 for vowel backness). In addition, the script returns a number of measures of Vowel Inherent Spectral Change. The measures of VISC for which results were analyzed are summarized in Table 2.[5]

---

[4] See (Eager 2015) for a discussion of the benefits of using the object window over the viewing window for automated measurements

[5] Additional measures returned by the script include F1-F2 Euclidean distance, time to formant target (as a percentage of vowel duration). We will not include measures that fit parametric curves to formant trajectories such as the Discrete Cosine Transform (DCT; e.g. Morrison 2013; Williams and Escudero 2014). According to Morrison (2013), there is no evidence that DCT models perform better than interval models with regard to vowel categorization.

Table 2. Summary of Acoustic measures of VISC

Measure	Description	Unit
Formant excursion	Total magnitude of formant movement over a selected interval	Bark
Formant slope	Mean rate of change for formant over a selected interval	Bark/100 msec
Formant stability	Proportion of total vowel duration that falls within 0.5 Bark of maximum or minimum value	Proportion (between 0 and 1)

A few comments concerning these measures are necessary at this time. Each of these measures is calculated on single formants. Thus, we will describe results in terms of Slope, Excursion, and Stability of F1 and F2 individually. Measures of F1-F2 Euclidean distance will not be presented (cf. Fox and Jacewicz 2009). This is because Euclidean distance makes use of absolute values, and as such gives a measure that is always positive. Since we are interested in both positive and negative formant movement, excursions based on single measures are described. In the case of Formant Slopes, the values of two formants could be combined into a single angle of formant trajectory in two dimensional space (e.g. Jin and Liu 2013). However, for clarity of presentation the Bark/100 msec unit for single formants will be used instead. Since 100 msec may be seen as a 'typical' vowel duration, and 1 Bark is an auditory critical band, the Bark/100 msec unit including both positive and negative values should be easier for the reader to interpret than a single positive number between 0 and 360 degrees.

The formant stability measure, developed by Schwartz (2010), allows for a characterization of entire vowels, rather than individual intervals, in terms of the 'purity' of vowel quality. Since interval measures treat portions of the vowel as separate units, it may be difficult to determine which and how many intervals are relevant for vowel categorization.[6] By contrast, the formant stability measure offers a useful, more holistic description of vowel purity, or lack of VISC. This is particularly useful to explore the hypothesis that languages differ systematically with regard to their degree of vowel purity.

Interval results presented here are based on an allotment of four intervals per vowel, each equal to 25% of the total vowel duration. An illustration of this partitioning is given in Figure 2, which shows a spectrogram of the FLEECE vowel segmented into four intervals. Also, since formant excursion measures resembled formant slope measures quite closely, to save space, only the slope results, and not the excursion measures, will be shown. Statistical results are based on a series of one-way ANOVAs with acoustic measures as dependent variable and participant group (Students vs. Advanced) as independent variable, with a .05 significance threshold.

---

[6]   This is an empirical question that we plan to take up in the future, employing discriminant analysis to evaluate the relative effects of particular intervals for vowel categorization.
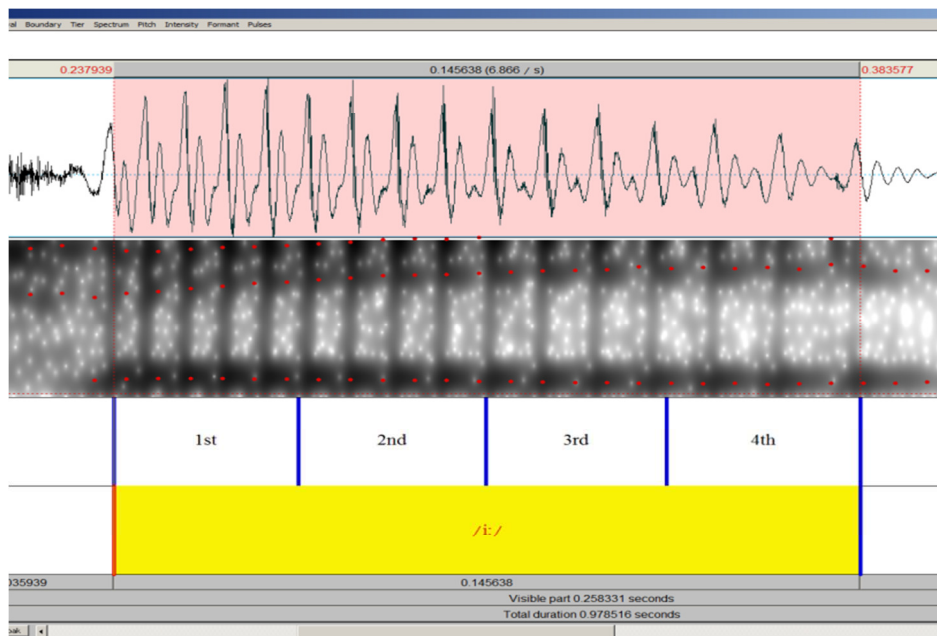
Figure 2. Illustration of vowel segmentation into component intervals

3.3. Results - FLEECE

The first set of results we will present comes from the production of the /iː/, for which we find a series of significant effects of Participant Group for both F1 and F2. F1 Slope measures are shown in Figure 3, in which significant differences in formant trajectories were observed for all intervals except the third (1st: F[1, 353]=10.54, p=.001; 2nd F[1, 353]=28.8, p<.001; 3rd: F[1, 353]=3.09, p=.08; F[1.353]=5.74, p=.017). As can be seen in the figure, the most dramatic difference between the two group is observed in the first interval, which the Teachers showed a falling F1 while the Students showed a rising F1. Note also that in the second and third intervals, the Students F1 slopes are flatter than those of the Teachers, which is suggestive of a more steady vowel quality.
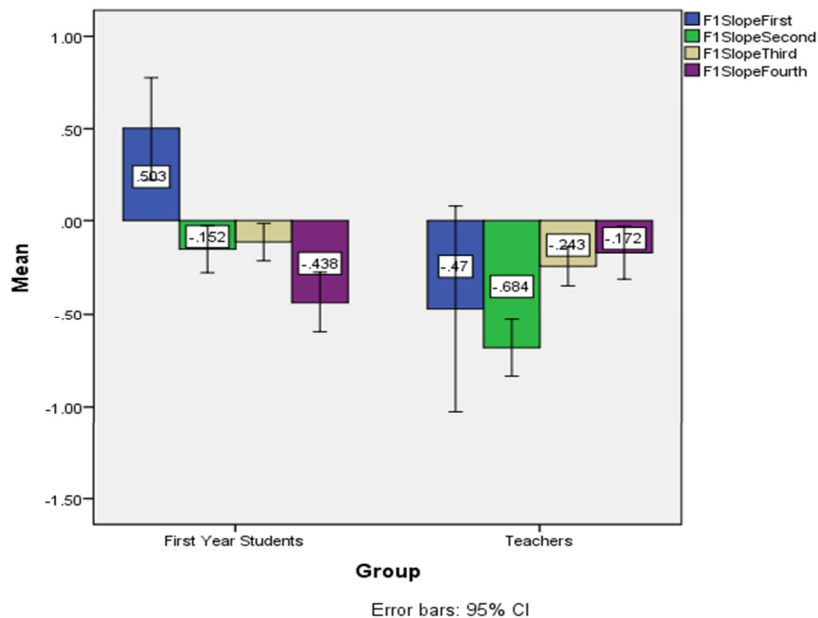
Figure 3. Formant slope measures for F1 of the FLEECE vowel (error bars represent 95% confidence intervals)

The relative steadiness of the F1 trajectory in the Students' productions is confirmed by the comparison of F1 Stability, shown in Figure 4 (see next page). On average 66% of Students' /i:/ durations were characterized by an F1 that fell within one-half Bark of the target, compared to 55% for the Teachers, $F[1, 353]=10.84$, $p=.001$.

With regard to F2 of the FLEECE vowel, significant differences between the groups were observed in every interval except for the third (1st: $F[1, 353]=13.4$, $p<.001$; 2nd: $F[1, 353]=8.43$, $p=.004$; 3rd: $F[1, 353]=1.63$, $p=.203$; 4th: $F[1, 353]=6.05$, $p=.014$). This is shown in Figure 5 (see next page). As with F1, the first half (two intervals) of the vowel are characterized by greater formant movement on the part of the Teachers, and the fourth interval shows greater movement by the Students.
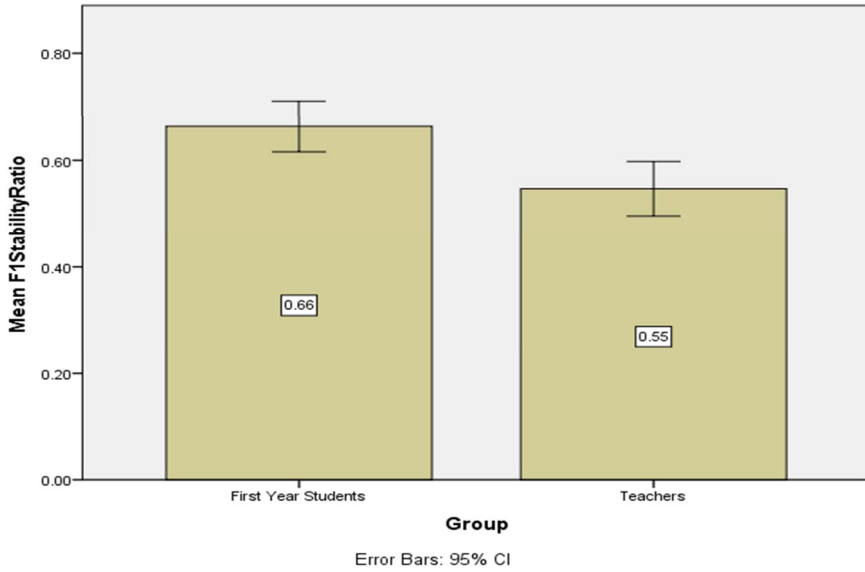
Figure 4. Formant stability for F1 of the FLEECE vowel, (error bars represent 95% confidence intervals)
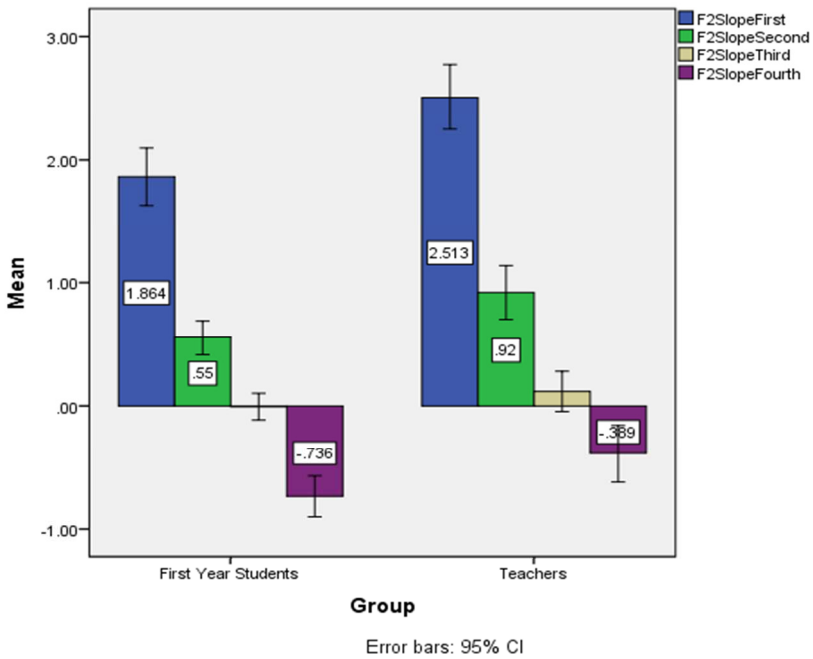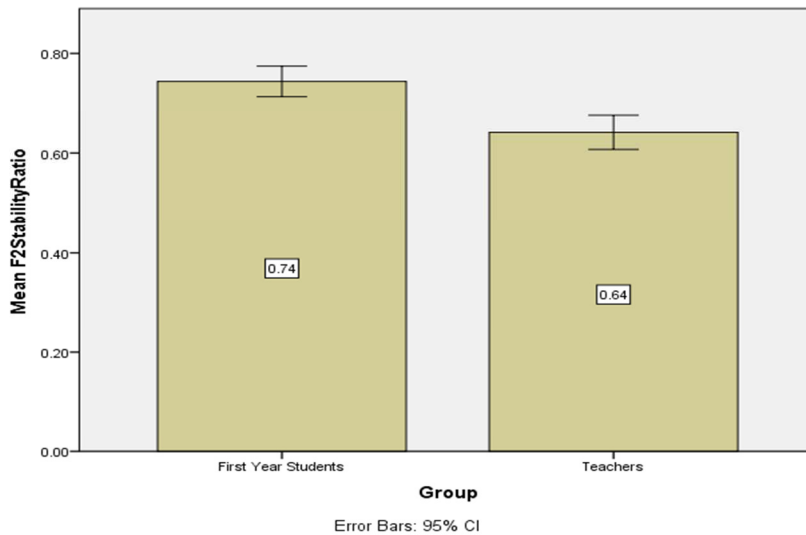

Figure 5. Formant slopes for F2 of the FLEECE vowel, (error bars represent 95% confidence intervals)

F2 stability measures for the FLEECE vowel are shown in Figure 6. The Students averaged 74% F2 stability while the Teachers averaged 64%, F[1, 353]=20.5, p<.001.

Figure 6. Formant stability for F2 of the FLEECE vowel, (error bars represent 95% confidence intervals)

3.4. Results – TRAP

In the case of the TRAP vowel, we will only present results from F1, since F2 did not show any significant effects of participant group. F1 slope measures for /æ/ are given in Figure 7 (see next page). Significant differences between the participant groups were observed in all intervals except the fourth (1st: F[1,322]=20.7, p<.001; 2nd: F[1, 322]=9.96, p=.002; 3rd: F[1,322]=34.9, p<.001; 4th: F[1, 322]=.652, p=.42). As with /i:/, the highest magnitude differences between groups were found in the first two intervals, for which the Teachers exhibited steeper F1 slopes.

This general pattern is also apparent in the F1 stability measures, shown in Figure 8 (see next page). The Students showed more stable F1 patterns than the Teachers (61% vs. 53%), F[1,322]=14.3, p<.001.
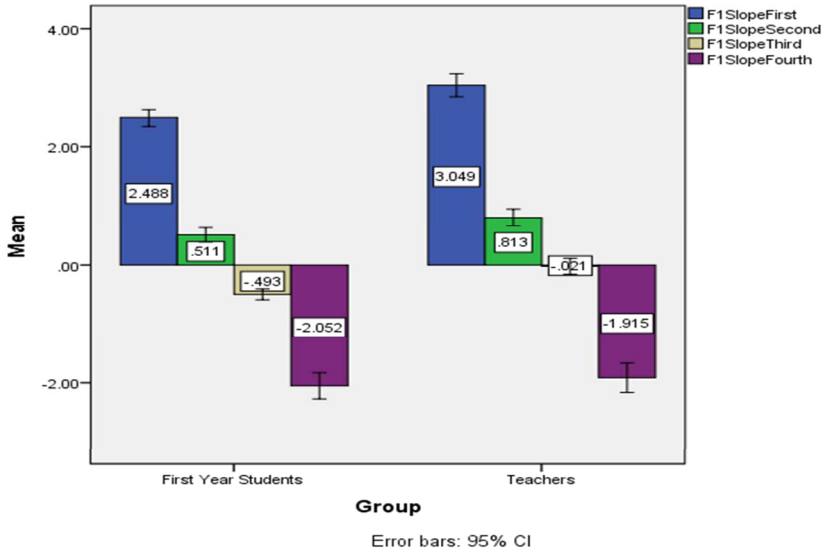
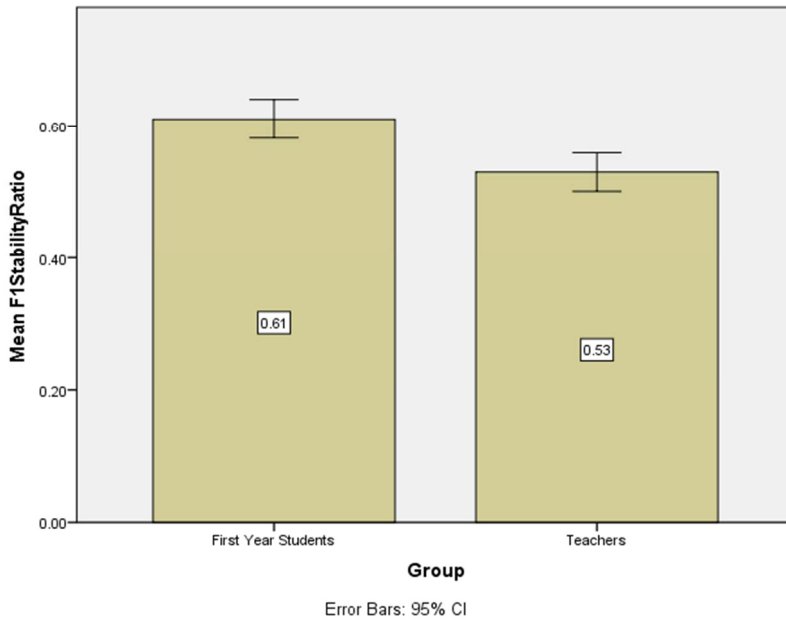**Figure 7**. Formant slopes for F1 of the TRAP vowel, (error bars represent 95% confidence intervals)

Figure 8. Formant stability for F1 of the TRAP vowel, (error bars represent 95% confidence intervals)

3.5. Discussion

Overall, the results of our study indicate that more proficient Polish users of English produce more robust patterns of VISC in pronouncing the TRAP and FLEECE vowels. Significant differences were observed in both F1 and F2 stability for the FLEECE vowel, and in F1 for the TRAP vowel. In all of these areas, the Students showed more stability while the Teachers showed more movement. Interestingly, it was primarily the first and second intervals that were responsible for the differences between groups. This fact suggests that formant movement in both the CV transition (0-25 interval) and the 25-50 interval is an important aspect of the teachers' realization of these vowels,[7] while the movement in the second half of the vowel is a less integral aspect of the representation of English vowels. Our future work will include additional vowels in the analysis, to facilitate description of dynamic patterns involving difficult vowel contrasts such as FLEECE-KIT and DRESS-TRAP.

The findings reported in this paper are compatible with the basic claim, outlined in Schwartz (2015), that a primary source of VISC may be found in co-articulatory effects of onset consonants. When CV transitions are produced more slowly, effects of co-articulation extend further into the vowel and may become reinterpreted as a feature of the vowel itself. Slower CV transitions may be expected due to the fact that English is generally categorized by a lower degree of supra-laryngeal muscular tension than Polish in its articulatory setting (Święciński 2004). Our future work will compare the effects of different places of articulation of both onset and coda consonants.

4. General Discussion – Visualizing VISC in L2 vowel production

To this point we have documented differences in the realization of the FLEECE and TRAP vowels in two groups of Polish users of English in terms of measures of spectral dynamics and formant stability. One useful thing about these metrics is that they allow for comparison across speakers, since formant slope and stability measures expressed in Bark capture phonetic properties that are independent of vocal tract size and anatomy. At the same time, however, one could argue that the pedagogical utility of such measures may be limited, since it is difficult to instruct students to produce a certain formant slope without explicit reference to targets in acoustic space. The role of spectral dynamics in pronunciation pedagogy must therefore be expressible in terms of the two-dimensional vowel representations that are more familiar to both teachers and students. In this connection, the role of VISC may be visualized as a series of static targets in different intervals of the vowel. It is not intuitively difficult to

---

[7]   One question that needs to be addressed in future work is the degree to which the teachers' productions resemble those of native speakers.

express to students that a vowel may be divided into a number of intervals, so teachers and students may examine static values in different portions of a vowel as a way of extrapolating more pervasive dynamic patterns.

Figure 9 shows mean F1 values (bark normalized into F1-f0 values) for the FLEECE vowel in each of the four intervals analyzed.[8] The red lines represent the Teachers' values, while the blue lines represent those of the Students. Significant differences were observed in the first interval and the third interval, but not in the second and fourth. Thus, in some sense, we may suggest that the Teachers have internalized a representation in which the low F1 target associated with a high vowel is reached relatively late, only in the third quarter of the vowel's duration. This pattern corresponds with diphthongal realizations of /iː/ that are the norm for Southern British English.
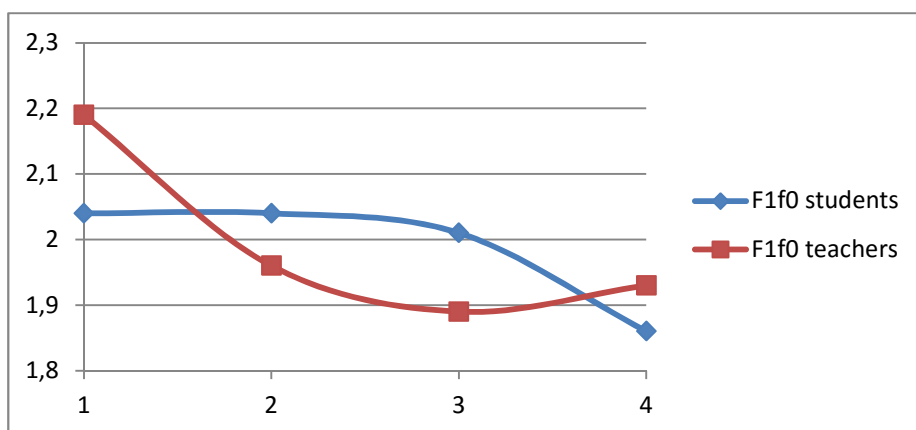
Figure 9. Mean F1-f0 measures per interval in the FLEECE vowel

This general pattern of diphthongization of FLEECE is also observable in mean values of F2, which is shown in Figure 10. In each of the intervals, the Students produced a higher bark-normalized F2 than the Teachers, however, the magnitude of this difference is much greater only in the first interval. Extrapolating the mean F2 values into formant trajectories reveals that the F2 of the Students is much more stable, as suggested by the Stability Ratio (Figure 6).

---

[8]   A reviewer has taken issue with the figures in this section, suggesting that they are 'results' and do not constitute discussion. However, since these figures represent static, rather than dynamic formant measures, they do not, strictly speaking constitute results for our analysis of VISC measures. Rather, they are provided as a way of visualizing VISC for pedagogical purposes, as suggested in the title of this section.
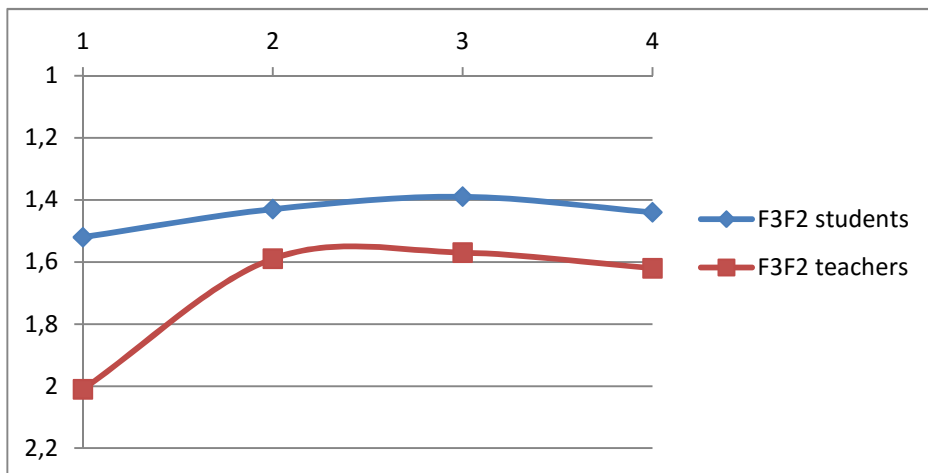
Figure 10. Mean F3-F2 measures for the FLEECE vowel

In Figure 11 (see next page) we see mean F1 values for each interval of the TRAP vowel. The interesting pattern to observe here is that in the first interval, the F1 measures between the groups do not differ – they only diverge from the second interval onward. The results shown in Figure 10 have implications for the fact that beginning learners often substitute Polish /e/ for the TRAP vowel. This error may result from Polish listeners' reliance on the initial portion of the vowel, with a fairly low F1 typical for /e/ and indicating that the low target position has not yet been reached. That is, since Polish vowels are typically pure in quality, targets are reached early in Polish, so listeners may base their percept in an earlier portion of the vowel. By contrast, greater VISC in English requires native listeners to tune in to the later portions of the vowel as well. This interpretation is also compatible with perception data presented by Jekiel (2010), and Schwartz et al. (2015).

The Mean F1 values for each vowel interval shown for the TRAP vowel provide an instructive example of the broader perspective that VISC can provide on the phonetics of vowel quality. In recent years, many studies have shown that in Southern British English, the target position for the TRAP vowel is becoming increasingly lower and retracted. For example, Hawkins and Midgley 2005 show that younger speakers produce /æ/ with a much lower F1 and Lower F2 than older generations. At first glance, the characterization of /e/-like quality of the first interval of the vowel might appear to be at odds with these descriptions of lowering and retraction. In this connection, however, it must be remembered that studies such as Hawkins and Midgley's were based on steady-state formant measurements taken at or near vowel midpoint. Thus, the claim that native-like tokens have an /e/-like first interval is not at all incompatible with descriptions of a lower and retracted vowel, based on formants measures near vowel midpoints. Indeed the F1-f0 measures in the 2nd and 3rd intervals in Figure

suggest a fairly significant degree of lowering in the Teachers' productions – it is only the first interval that is /e/-like. What Figure 11 suggests is that cross-language vowel perception may be based on different portions of a vowel. Static formant measures based on steady-state do not allow us to describe such patterns.
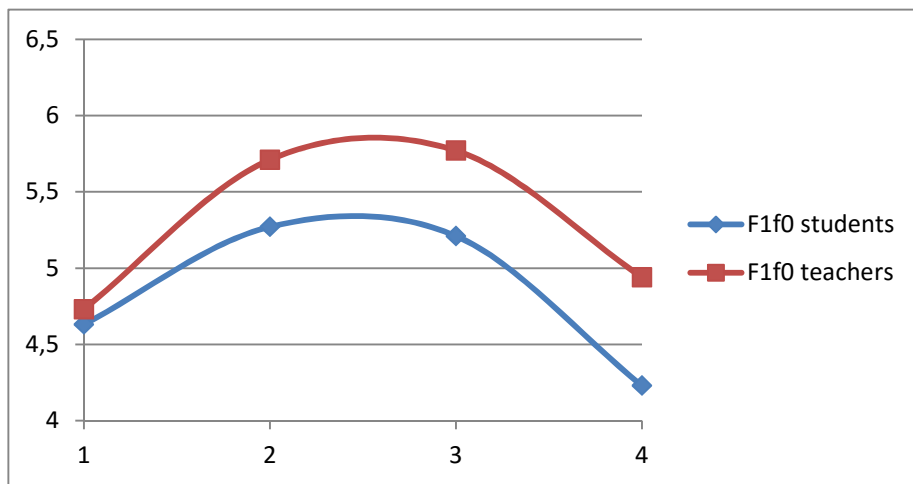

Figure 11. Mean F1-f0 values for the TRAP vowel

To conclude, this paper has provided data on the production of Vowel Inherent Spectral Change by Polish users of English. The group of teachers, with professional-level proficiency in English, produced more dramatic patterns of spectral movement in the FLEECE and KIT vowels than the group of first year students. These results may be interpreted in terms of phonological parameter settings that are encoded with the representations of the Onset Prominence framework. According to the theory, consonant representations in English encroach upon the acoustic space of the vowel, opening the door to greater consonant-vowel interactions that result in formant movement, while in Polish these effects are limited. As a result, greater proficiency in English entails the acquisition of these dynamic patterns, as we have observed in the data presented in this paper.

References

Aperliński, G. and G. Schwartz. 2015. Release bursts vs. formant transitions in Polish stop place perception. In The Scottish Consortium for ICPhS 18 (ed.), Proceedings of the 18th International Congress of Phonetic Sciences.

Bertinetto, P. 1989. Reflections on the dichotomy stress vs. syllable-timing. *Revue de Phonetique Appliquee* 91-93. 99-130.

Best, C. 1995. A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171-204. Timonium, MD: York Press.

Best, C. T. and M. D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro and O.-S. Bohn (eds.), *Second language speech learning – the role of language experience in speech perception and production,* 13-34. Amsterdam: John Benjamins

Bohn, O.-S. 1995. Cross language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues,* 279-304. Timonium, MD: York Press.

Bohn, O.-S. and J. Flege. 1997. Perception and production of a new vowel category by second-language learners. In A. James and J. Leather (eds.), *Second-language speech: Structure and process,* 53-74. Berlin: Walter de Gruyter.

Docherty, G. 2010. Phonological innovation in contemporary spoken British English. In A. Kirkpatrick (ed.), *The Routledge Handbook of World Englishes,* 59-75. Routledge, London.

Donegan, P. and D. Stampe. 1983. Rhythm and the holistic organization of language structure. In *Papers from the Parasession on the Interplay of Phonology, Morphology, and Syntax,* 337-353. Chicago: CLS 19.

Draxler, C. and K. Jänsch. 2015. SpeechRecorder v. 2.X.X. [Software]. Available from: http://www.bas.unimuenchen.de/Bas/software/speechrecorder/

Dukiewicz, L. and I. Sawicka 1995. *Gramatyka współczesnego języka polskiego – fonetyka I fonologia* [Grammar of modern Polish – phonetics and phonology]. Krakow: Wydawnictwo Instytutu Języka Polskiego PAN.

Eager, C. 2015. Automated voicing analysis in Praat: Statistically equivalent to manual segmentation. In The Scottish Consortium for ICPhS 18 (ed.), Proceedings of the 18th International Congress of Phonetic Sciences

Escudero, P. and P. Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. Studies in Second Language Acquisition 26. 551-585.

Fox, R. A. and E. Jacewicz. 2009. Cross-dialectal variation in formant dynamics of American English vowels. Journal of the Acoustical Society of America 126. 2603–2618. doi:10.1121/1.3212921

Flege, J. E. 1987. The production of 'new' and 'similar' phones in a foreign language: Evidence for equivalence classification. Journal of Phonetics 15. 47-65.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues,* 233-277. Timonium, MD: York Press.

Gonet, W., Szpyra-Kozłowska, J. and R. Święciński. 2010. Clashes with ashes. In E. Waniek-Klimczak, E. (ed.), *Issues in Accents of English 2*, 213-232. Newcastle upon Tyne: Cambridge Scholars Publishing.

Hawkins, S. and J. Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. Journal of the International Phonetic Association 35/2. 183-199.

Hillenbrand, J. 2013. Static and dynamic approaches to vowel perception. In G. Morrison and P. Assmann (eds.), *Vowel inherent spectral change*, 9-30. Berlin: Springer.

Jekiel, M. 2010. Dynamic information for Polish and English vowels in syllable onsets and offsets. Unpublished B.A. thesis, Adam Mickiewicz University in Poznań.

Jenkins, J. J. and W. Strange. 1999. Perception of dynamic information for vowels in syllable onsets and offsets. Perception and Psychophysics 61. 1200–1210.

Jin, S. H. and C. Liu. 2013. The vowel inherent spectral change of English vowels spoken by native and non-native speakers. Journal of the Acoustical Society of America 133 (5). 363-369. Available from: http://dx.doi.org/10.1121/1.4798620.

Kaźmierski, K. 2014. Are some languages more prone to vowel shifting than others? Paper delivered at the 47th Annual Meeting of the Societas Linguistica Europaea September 11-14 2014, Poznań.

Labov, W., S. Ash and C. Boberg. 2006. *Atlas of North American English: Phonetics, Phonology, and Sound Change.* Berlin: Mouton de Gruyter.

Lass, N. 1996. *Principles of Experimental Phonetics*. St. Louis: Mosby.

Laver, J. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Ladefoged, P., 1999. *American English. Handbook of the International Phonetic Association*, 41-44. Cambridge: Cambridge University Press.

Lindblom, B. 1990. Explaining phonetic variation: a sketch of the HandH theory. In W. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modeling*, 403-439. The Netherlands: Kluver Academic.

Kaźmierski, K. 2015. *Vowel-shifting in the English language*. Berlin: Mouton de Gruyter.

Malisz, Z. 2013. Speech rhythm variability in Polish and English – a study of variability in rhythmic levels. PhD dissertation, Adam Mickiewicz University in Poznań

Morrison, G. 2013. Theories of Vowel Inherent Spectral Change. In G. Morrison and P. Assmann (eds.), *Vowel inherent spectral change,* 31-48. Berlin: Springer.

Morrison, G and P. Assmann (eds.). 2013. *Vowel inherent spectral change*. Berlin: Springer.

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

Ramus, F., M. Nespor and J. Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73. 265–292.

Rogers, C. L., M. Glasbrenner, T. DeMasi and M. Bianchi. 2013. Vowel inherent spectral change and the second language learner. In G. S. Morrison and P. Assmann (eds.), *Vowel Inherent Spectral Change, Modern Acoustics and Signal Processing,* 231-259. Berlin/Heidelberg: Springer-Verlag.

Rojczyk, A. 2011. Overreliance on duration in nonnative vowel production and perception: The within lax vowel category contrast. In M. Wrembel, M. Kul and K. Dziubalska-Kołaczyk (eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010*, vol. 2, 239-249. Bern: Peter Lang.

Schwartz, G. 2010. Rhythm and vowel quality in accents of English. Research in Language 8. 135-147.

Schwartz, G. 2013. A representational parameter for onsetless syllables. *Journal of Linguistics* 49(3). 613-646.

Schwartz, G. 2015. Vowel dynamics for Polish learners of English. In: E. Waniek-Klimczak and M. Pawlak (eds.), *Teaching and Researching the Pronunciation of English - Studies in Honour of Włodzimierz Sobkowiak,* 205-217. Berlin: Springer.

Schwartz, G. 2016. On the evolution of prosodic boundaries – parameter settings for Polish and English. Lingua 171. 37-73.

Schwartz, G. and G. Aperliński. 2014. The phonology of CV transitions. In E. Cyran and J. Szpyra-Kozłowska (eds.), *Crossing Phonetics-Phonology Lines*, 277-298. Newcastle: Cambridge Scholars Publishing.

Schwartz, G., Aperliński, G., Jekiel, M. and K. Malarski 2015. Spectral dynamics in L1 and L2 vowel perception. Paper presented at the 9th International Conference on Native and Non-native Accents of English, Accents 2015, University of Łódź.

Schwartz, G., Balas, A. and A. Rojczyk. 2014. Stop release in Polish English – implications for prosodic constituency. Research in Language 12(2). 131-144.

Strange, W. 1989. Evolving theories of vowel perception. Journal of the Acoustical Society of America 85. 2081- 2087.

Strange, W., Jenkins, J. and T. Johnson. 1983. Dynamic specification of coarticulated vowels. Journal of the Acoustical Society of America 34. 695-705.

Syrdal, A. and H. S. Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustic Society of America* 79(4). 1086-1100.

Święciński, R. 2004. Articulatory setting in Polish and its implications for teaching English pronunciation to Poles. In W. Sobkowiak and E. Waniek-Klimczak (eds.), *Dydaktyka fonetyki języka obcego. Zeszyt Naukowy Instytutu Neofilologii Państwowej Wyższej Szkoły Zawodowej w Koninie nr 3*, 141-150. Konin: Wydawnictwo PWSZ w Koninie.

Wagner, P. 2007. Visualising levels of rhythmic organization. *Proceedings of ICPhS XVI. Saarbrucken.*

Weckwerth J. 2011. English TRAP vowel in advanced Polish learners: Variation and system typology. In W.-S. Lee and E. Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 2110-2113. Hong Kong: City University of Hong Kong.

Williams, D. and P. Escudero. 2014. A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. Journal of the Acoustical Society of America 136(5). 2751-2761.

White, L. and S. L. Mattys. 2007. Calibrating rhythm – first and second language studies. *Journal of Phonetics* 35. 501-522.