*Poznań Studies in Contemporary Linguistics 54(2), 2018, pp. 307–332*

*doi: 10.1515/psicl-2018–0007*

# POLISH LISTENERS' PERCEPTION OF VOWEL INHERENT SPECTRAL CHANGE IN L2 ENGLISH

GEOFFREY SCHWARTZ* AND JERZY DZIERLA
*Adam Mickiewicz University, Poznań*
**geoff@wa.amu.edu.pl*

ABSTRACT

This paper describes a perception experiment with Polish listeners involving vowel inherent spectral change (VISC) in L2 English. A forced-choice rhyming task employing the Silent Center (SC) paradigm revealed relatively uniform effects of stimulus type (SC, Initial, Middle, Final) on accuracy across two proficiency groups, despite greater overall accuracy on the part of the more proficient users. Analysis of individual vowel pairs used in the rhyming trials revealed some effects of proficiency on the degree to which formant movement in the stimuli affected identification accuracy. This research contributes to the relatively sparse literature on VISC in L2 acquisition. Phonological considerations underlying the degree of VISC in Polish and English are also discussed.

KEYWORDS: Vowel inherent spectral change; L2 speech perception; phonetics and phonology.

## 1. Introduction

Most research into second language (L2) vowel acquisition is driven by cross-language comparisons of vowel inventories, which form the basis for predictions with regard to expected difficulties for L2 acquisition. In this endeavor, particular attention has been paid to cases in which the L2 features contrasts lacking in L1, often between vowels which are in relatively close proximity in F1–F2 vowel space. When the target language for acquisition is English, learners are faced with a number of notorious pairs, such as *sheep–ship*, *look–Luke*, *men–man*, and *lock–luck*, which have attracted the attention of L2 speech researchers.

In probing these issues, investigators have compared the perceptual weight of different types of acoustic cues used by listeners. Most frequently, these comparisons have examined the role of vowel duration as opposed to formant targets in F1–F2 space. One interesting finding is that L2 learners from L1s without vowel duration contrasts may use duration cues in discriminating L2 contrasts. For example, both Bohn (1995) and Escudero and Boersma (2004) describe findings by which L1 Spanish speakers place more weight on duration cues while native speakers attend more to spectral cues in distinguishing the vowels in *beat* and *bit*. Rojczyk (2011) made a similar observation in an experiment on L1 Polish speakers' discrimination of English /æ/ from /ʌ/. Despite the fact that Polish has no duration contrasts, learners noticed the longer duration of /æ/ and use it to discriminate the vowel from /ʌ/. These findings are compatible with what Bohn (1995) has described as a type of perceptual "desensitization" by L2 learners. The claim is that when a new L2 vowel sound is in close spectral proximity to an L1 sound, listeners are "desensitized" to its spectral details, promoting duration as the best available cue for learner discrimination. Bohn's desensitization hypothesis is closely related to the postulate of Flege's Speech Learning Model (SLM) by which L2 sounds that are phonetically similar to L1 sounds are subject to equivalence classification (Flege 1987), hindering acquisition.

While this and similar research is invaluable for our understanding of L2 speech perception, they avoid a more general question: how is it that contrasts between spectrally similar vowels arise in the first place? In other words, shouldn't children learning English as an L1 also become desensitized to spectral similarity and start merging difficult contrasts? One relatively new current in phonetics research may provide an answer to these questions. Vowel Inherent Spectral Change (VISC; e.g. Morrison and Assmann 2013), i.e. changes in formant frequencies over the time course of a vowel, is becoming an increasingly prominent element of descriptions of L1 English vowel systems. Traditionally descriptions of vowels are based on static target positions in a two-dimensional acoustic space, with no representation of time apart from diacritics denoting length or clipping. By looking at VISC, we may see how two vowels that have similar "target" positions in fact may have greatly different dynamic properties. For example, it has been observed that the contrasts between long and short (lax and tense) high vowels in English, which are notoriously difficult for L2 learners, is based in large measure on the direction of formant movement. Tense vowels tend to move more toward the periphery of F1–F2 space, while lax vowels tend to show movement toward the center (Nearey and Assmann 1986; Nearey 2013).

VISC also has implications for L2 speech acquisition research, which has been slow to incorporate spectral dynamics into descriptive and empirical work on the production and perception of L2 vowels. This fact is somewhat surprising, since textbook descriptions of English vowels (e.g. Cruttenden 2001; Collins and Mees 2009) often mention diphthongization as a characteristic feature of the vowel system of many native varieties. Nevertheless, VISC has yet to find its way into the mainstream of L2 speech research. The present paper is part of a larger project which looks at the acquisition of English VISC by L1 Polish learners. Since Polish is a language with a simple vowel system and relatively stable vowel quality (Dutkiewicz and Sawicka 1995), our goal is to consider the dynamic vs. static aspects of L2 English and L1 Polish vowels as a new parameter for calculating cross-language phonetic "similarity", which of course is a crucial concept for current models of L2 speech acquisition (Flege 1995; Best 1995; Best and Tyler 2007).

In this paper, we describe a perceptual experiment in which we examine how formant movement affects learner identification accuracy at two different levels of proficiency. As a corollary, our experimental design, in which different portions of a vowel's duration are used as stimuli, will allow us to say something about the time course of L2 English vowel perception by Polish listeners. In other words, we consider the following questions: what is the relative perceptual weight of various portions of an English vowel (the beginning, the middle, the end), and how does this change as a function of L2 proficiency? The rest of this paper will proceed as follows. Section 2 will provide some background information on VISC. Section 3 describes the experiment itself. Section 4 concludes with a brief discussion of the phonological underpinnings of the hypothesis investigated in this study. It is shown how dynamic effects on vowel quality may emerge on a language-specific basis in accordance with the representational settings in the Onset Prominence framework (Schwartz 2013, 2016).

## 2. Background on VISC

The origins of VISC research date back to acoustic experiments in the 1950s and 1960s. In one study, Peterson and Lehiste (1960) looked at a number of acoustic aspects of American English vowels. In addition to patterns of inherent duration, they described vowel-based differences in formant trajectories, measuring the duration of CV and VC transitions, as well as the duration of the quasi-steady-states in vowel nuclei. In later work (e.g. Nearey and Assmann 1986), these observations were documented in more detail, and certain generalizations

were formulated with regard to formant movement, primarily in North American English. Notably, it was observed that so-called "tense" vowels tend to show movement toward the periphery of the acoustic vowel space, while "lax" vowels are characterized by movement toward the center (see e.g. Nearey 2013: 52–54). On the basis of this work, the term Vowel Inherent Spectral Change (VISC) was coined, and VISC was hypothesized to be a truly inherent aspect of the vowels of North American English. To support this hypothesis, researchers often employed discriminant analyses, establishing VISC, in addition to static formant targets, as a significant predictor of vowel identity.

Early studies suggested a connection between VISC and the effects of neighboring consonants. In several early studies of vowel production, authors (e.g. Lindblom 1963 for Swedish; Stevens and House 1963 for American English) documented "target undershoot" in CVC contexts. Under the influence of neighboring consonants, canonical formant targets associated with vowels produced in isolation very often are not reached. The target undershoot problem became a focus for research into vowel perception, which asked how identification could remain constant even when the acoustics of a given vowel showed a great deal of consonant-induced variability. This issue was investigated in a current of experimental research in the 1970s and 1980s, which was carried out primarily in North America (for a review, see Strange 1989). A number of studies found that consonant-induced co-articulation did not hinder vowel identification. Indeed, throughout these studies, American English vowels produced in isolation were never identified more accurately than those embedded in CVC contexts, and in some cases it was the co-articulated vowels that were identified more accurately.

If VISC indeed originates in phonetic interaction between vowels and neighboring consonants (see Hillenbrand et al. 2001), we may consider its diachronic development in terms of a listener-oriented view of phonology (Ohala 1981; Blevins 2004) in which acoustic/perceptual ambiguities in the speech signal may lead to the reorganization of phonological specifications. We know from research into speech perception that formant transitions on vowels are used by listeners to identify consonant place of articulation (e.g. Wright 2004). Assuming that these transitions occupy the first (and last in the case of VC transitions) 20–25% of a vowel's duration – many vowel studies exclude the 0–20% and 80-100% intervals of a vowel in order to "reduce the effects of formant transitions associated with flanking consonants" (Williams and Escudero 2014: 2754) – we may expect consequences when the transitions are produced more slowly. Slower formant transitions extend further into a vowel's duration, to vowel midpoint or even beyond. When these dynamic formant patterns occupy

more of a vowel's duration, we should expect listeners to start interpreting that movement as a feature inherent to the vowel itself, rather than as a co-articulatory effect of the consonants. In other words, rather than formant transitions being used to perceptually "reconstruct" consonants (Ohala 1981), VISC becomes phonologized as part of canonical vowel representations.[1] We will return shortly to the issue of the role of consonantal context in the development of VISC. First, however, it is necessary to provide a more complete and up-to-date picture of the current state of VISC research.

In one current of research, experimenters have focused on production, employing VISC to refine acoustic descriptions of sociolinguistic variation within and across dialects of English. For example, one set of studies described in Fox and Jacewicz (2009) and Jacewicz and Fox (2013), found that younger North American speakers, both in Northern and Southern dialect areas, show a lesser amount of formant movement than older speakers, measured in terms of the sum of Euclidean distances over four vowel-internal intervals. One possible interpretation of this finding is that while formant movement plays a role in diachronic vowel shifts, after such shifts are established in younger generations, the new vowel identities become regularized, leading to more stable vowel quality. Williams and Escudero (2014) compared vowel qualities in Southern British English and Sheffield English. They found that adding formant trajectories to discriminant analyses based on mean formant values, increased classification accuracy for both dialects. In a similar vein, Elvin et al. (2016) looked at vowel formant dynamics in the dialect of English spoken in Western Sydney, Australia, and found an important role of VISC in the classification of vowel identities of both diphthongs and nominal monophthongs. Finally, Williams et al. (2015) is one of just a few cross-language VISC studies, comparing dynamic formant patterns in British English and Dutch.

L2 speech production has been the subject of only a small number of VISC studies. Jin and Liu (2013) compare native American English speakers to L1 Chinese and Korean learners. They found that the L1 Chinese speakers exhibited the greatest degree of VISC among the three groups of speakers, exceeding even that produced by the L1 English speakers. Research carried out by Rogers et al. (2013) compares L1 American English speakers with bilingual Spanish and English speakers in South Florida. These authors found that native speakers

[1] Of course, different consonant places of articulation are associated with different formant trajectories. In this regard a further hypothesis may be formulated that warrants testing. Namely, trajectories associated with the most common place of articulation, typically coronal, may be extended analogically throughout the vowel system.

and early bilinguals produced very similar formant trajectories, in which VISC served to distinguish vowels that are located in close proximity in the vowel space, while late learners of English produced a lot of acoustic overlap, particularly in the case of the front vowels over 3 different measurement points (Rogers et al 2013: 248, Fig. 6). Some research may be found investigating how Polish learners of English, whose L1 is characterized by a relatively stable vowel system, acquire the dynamic properties of English vowels. Schwartz et al. (2016a) compared production of the FLEECE and TRAP vowels with labial or coronal onsets by B1-level Polish learners of English with L1 Polish speakers with C2-level proficiency in English,[2] and found that more proficient users used a greater degree of formant dynamics, particularly in the first half of the vowel's duration, while the less proficient learners produced more stable formant trajectories.

Research into VISC perception in English has been carried out primarily in North America, and has attempted to identify which portion or portions of a vowel's duration comprises what may be thought of as the perceptual identity of English vowels. Motivated by the target undershoot problem, a number of studies compared identification rates of vowels co-articulated in CVC contexts with those produced in isolation (for reviews, see Strange 1989 and Hillenbrand 2013). Throughout these studies, American English vowels produced in isolation were never identified more accurately than those embedded in CVC contexts, and in some cases it was the co-articulated vowels that were identified more accurately. These findings led to the formulation of a hypothesis that static formant targets in a two-dimensional space were insufficient in describing the perceptual identity of American English vowels. Rather, in the "dynamic specification" approach (Strange, 1989), formant trajectories over the duration of the vowel also provide listeners with crucial cues for vowel perception.

To test this hypothesis, an experimental paradigm was developed in which naturally produced stimuli were altered by silencing various parts of a vowel's duration, allowing researchers to investigate in a controlled fashion the role of formant dynamics in vowel identification. In one such stimulus condition, referred to as the Silent Center condition (SC; e.g. Strange et al. 1983), the central quasi-steady-state portion of the vowel is silenced, leaving listeners to identify vowels on the basis of CV and VC transitions. Silent Center tokens are compared for perception accuracy with tokens in which central portion of the vowel is included, or others in which only the CV or VC transitions are included, or unmodified tokens. A consistent finding in these experiments with North Ameri-

[2] Proficiency levels in accordance with the Common European Framework of Reference (CEFR).

can listeners was that the SC tokens were identified most accurately of all the modified stimuli, with error rates often not significantly higher than unmodified tokens (Strange, 1989; Jenkins and Strange, 1999). Other stimulus conditions, especially those based only on the CV or VC transition, but also those based only on the portion near the vowel midpoint, induced higher error rates from North American listeners.

The SC studies suggest a significant role for formant movement induced by CV and VC transitions in L1 English vowel perception. In languages with less VISC, however, we should expect different results. In a study with direct bearing on the present experiment, Schwartz et al. (2016b) employed the Silent Center paradigm to look at Polish vowel perception (see also Jekiel 2010). They found no effects of stimulus Type (Initial, Middle, Silent Center, Final) on identification accuracy. Polish listeners were highly accurate regardless of which portion of the vowel they heard. In the present study, we focus on the acquisition of English VISC by Polish learners, employing the SC paradigm to compare dynamic specification effects on vowel identification at two different levels of proficiency.

## 3. Perception experiment

This section will present the perception experiment carried out with Polish learners of English. The experiment employed what we believe to be a new method, a forced-choice rhyming task instead of simple identification task, with five different types of stimuli designed to investigate learners sensitivity to formant movement. More details about the stimuli and the procedure will be provided in the sections that follow.

### 3.1. Participants

There were two groups of participants in the experiment. The first group was comprised of 39 B1-level Polish students of English at the beginning of their first year of studies (Students). The second group was composed of 20 Ph.D. students and lecturers at the Faculty of English, all with C2-level proficiency (Teachers). The Students' group received points toward course credit in return for the participation. The Teachers' group was not compensated in any way for their participation.

### 3.2. Stimuli

The stimuli were created from recordings of a single male native speaker of Southern British English, who produced a series of carrier phrases designed to isolate individual vowels of English. For example, the phrase to isolate the vowel /ɪ/ read *In gick and gicka, we have /ɪ/*. The underlined second word in the phrase was used for the stimuli (cf. Williams and Escudero 2014). Stimuli were produced from tokens of the following vowels: /iː/, /ɪ/, /ɛ/, /æ/, /uː/, /ʊ/, /ʌ/, and /ɒ/ (Keywords: FLEECE, KIT, DRESS, TRAP, GOOSE, FOOT, STRUT, LOT).[3] The base recordings for stimuli were all CVC words in which the onset and coda had the same place of articulation, either labial, coronal, or dorsal. In each case the onset was lenis (/b d g/), while the coda was fortis (/p t k/). This was done in order to produce stimuli in which the inherent differences in duration between long and short English vowels are less dramatic due to pre-fortis clipping.

The recordings were edited in Praat to produce the stimulus conditions summarized in Table 1, inspired by "Dynamic Specification" experiments described in Hilllenbrand (2013). The proportions of the vowels used for each stimulus type were arrived at by translating the most conservative manipulations (i.e. those with the largest portions of the vowels included) found in Jenkins and Strange's (1999) stimuli, which were based on the number of pitch periods, and converting them into percentages. Our initial auditory evaluation of the stimuli revealed that items including 3 or fewer pitch periods in the Initial, Middle and Final conditions were extremely difficult to identify. For this reason we included the larger portions. The numbers in the table were used as a base. The actually percentages in the stimuli varied slightly (1–2%) because of the editing procedure involved in stimulus production, in which cuts were made at the nearest zero crossings on the waveform to the calculated percentage-based points.

In order to characterize the effects of formant movement on the listener responses, it is necessary to describe the acoustics of the stimuli in some detail. Figures 1–3 present formant tracks of the Base items in the labial, coronal, and dorsal contexts, respectively. In the figures the entire vowel duration is plotted, and the dots denoting F1–F2 formant values increase in size over the time course of the vowel.

[3] These eight vowels were selected since they form four pairs from which forced-choice perception trials may be constructed.

Table 1. Summary of stimulus types for rhyming experiment.

| Stimulus Type | Description | Notes |
|---|---|---|
| Base | Original unaltered token | |
| Middle | The central 30% of the original vowel duration | Preceded and followed by silences of 20% of the vowel duration |
| Initial | First 35% of vowel | Followed by a silence equal to 50% of vowel duration |
| Final | Last 35% of vowel | Preceded by silence equal to 50% of vowel duration |
| Silent Center (SC) | First and last 20% of vowel | Silent center equal to 50% of vowel duration |

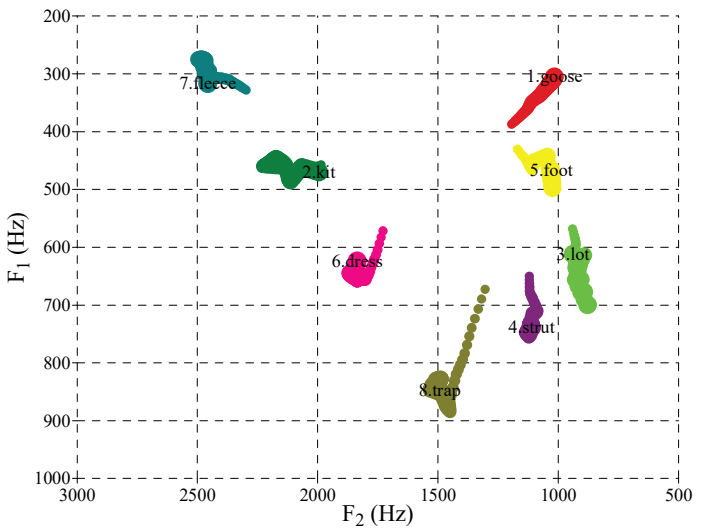


Figure 1. Formant tracks of base items used for stimuli – labial context. (Colour online.)

There are a number of aspects of the stimuli worth commenting on. First, it is apparent that there is a significantly lesser magnitude of formant movement in the labial context than in the coronal and dorsal contexts. This presumably reflects the independence of the lips from the tongue body as an articulator – in

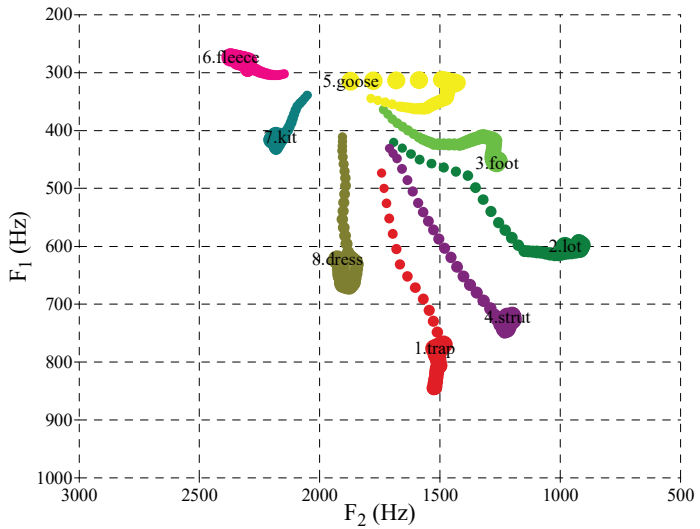


Figure 2. Formant tracks of base items used for stimuli – coronal context. (Colour online.)

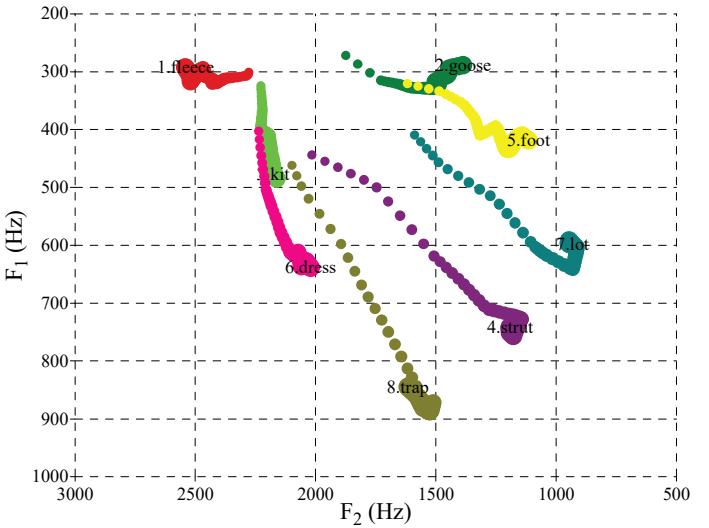


Figure 3. Formant tracks of base items used for stimuli – dorsal context. (Colour online.)

the labial context the tongue body is able to get a "head start" on its excursion to the target position for the vowel, leading to less formant movement. Also striking in the coronal context is the compactness of the vowel space in the early portion of the vowel (the smaller dots in the comet trails).

Table 2 summarizes the degree of formant movement for each stimulus type in each context collapsed across vowel qualities (for "static" formant information, see Appendix 1). The numbers reflect mean F1–F2 Euclidean distances measured in Bark for each stimulus condition and each consonantal context.

| | Labial | Coronal | Dorsal |
|---|---|---|---|
| Base | 1.19 | 2.51 | 2.61 |
| Initial | 0.63 | 1.59 | 1.68 |
| Middle | 0.31 | 0.55 | 0.61 |
| Final | 0.25 | 0.37 | 0.32 |
| SC | 0.69 | 1.23 | 1.42 |

Table 2. Summary of degree of formant excursion for each stimulus type and context. Numbers represent F1–F2 Euclidean distances measured in Bark.

In Table 2 we see that the Initial intervals have significantly more formant movement than the Middle and Final intervals. This movement may of course be attributed to the CV transition. Note also the minimal amount of movement in the Final items, which may be explained in terms of the fact that the codas were all fortis. It is likely that pre-fortis clipping is responsible for shortening the vowel, thus reducing formant movement in the VC transition. Also worth noting, and reflecting the measures from Table 2, is the smaller degree of movement in the labial CV transitions than in the coronal and dorsal transitions.

The degree of formant movement in different portions of the stimulus vowels may aid in the formulation of predictions for our study. If we assume that a greater degree of movement in a given interval should render identification more difficult, it may be hypothesized that Initial condition should be identified less accurately than the other conditions containing more stable portions of the vowel. Further, under this assumption we would expect items in labial contexts to be identified more accurately. We might also expect interactions among con-

text, stimulus type and proficiency level under the hypothesis that more proficient learners are less affected by VISC in the stimuli.

### 3.3. Procedure

Two primary considerations played a role in the decision to use a rhyming task, instead of a simple identification task. The first was an attempt to minimize biases that may come about from lexical frequency effects. Listeners may be biased toward more common words, and it is difficult to find forced-choice pairs of equal frequency. Thus, the rhyming words were all common and familiar items (see Appendix 2), and the experimenters made sure participants were familiar with them before the experiment started. In addition, for some consonantal contexts it is difficult to find actual words that could be used for identification. The rhyming task was designed to address this issue – it was less of a problem to find common words that simply rhyme with the stimuli, instead of perfect minimal pairs.

The experiment was carried out using E-Prime in the Language and Cognition Laboratory at the Faculty of English, Adam Mickiewicz University in Poznan. Each trial consisted of a single audio stimulus accompanied by two choices displayed on the screen. Upon hearing the audio stimulus, listeners were asked to identify which of the two words on the screen the recording rhymed with. In each trial the correct response was paired with a distractor from what might be considered the closest competitor. Thus, for FLEECE words the distractor contained the KIT vowel. DRESS and TRAP were paired together, as were GOOSE–FOOT and LOT–STRUT. The two rhyming choices were displayed on the screen for 500 milliseconds before the audio stimulus began playing. Keyboard input recorded the response, and advanced the experiment to the next trial. The rhyming choices were also paired with regard to whether the correct choice was displayed on the left side of the screen or the right side of the screen. A total of 240 trials (8 vowels × 5 stimulus types × 2 right-left pairings × 3 places of articulation) was divided into two blocks of 120.

### 3.4. Analysis

Statistical analyses were based on a total of 14,160 responses (59 total participants × 240 trials), of 9360 were collected from the Students group and 4800 from the Teachers' group. Analyses were performed with the SPSS statistical

software (IBM corporation 2013). The proportion of correct responses (accuracy) was the primary dependent variable of interest. Generalized linear mixed effects models with a logit transform to the binary target variable of accuracy included Stimulus Type and Group, and Context as fixed effects and Participant as a random effect. In the first, global analysis collapsed across vowel pairings, vowel Pair was included as a random factor. Subsequent analyses were based on subsets of the data corresponding to the four vowel pairs tested in the trials (FLEECE–KIT; DRESS–TRAP; GOOSE–FOOT; LOT–STRUT) for each group separately.

### 3.5. Results

We start by looking at the overall results collapsed across all vowel qualities, after which we shall investigate accuracy for each contrasting pair used in the experimental trials.

The overall accuracy rate was 72.8% (70.3% for the Students, 78.6% for the Teachers), which ranged from a minimum of 69.7% for the Initial tokens to 74.5% for the Final items. A generalized linear mixed model with a logit transform to the binary target variable of accuracy, and participant and vowel as random factors, revealed significant effects of participant group and stimulus type. The Teachers were more accurate overall, while the Initial tokens were identified least accurately. There was no significant effect of Context. The significant effects are shown in Table 3, and summarized graphically in Figure 4. The model revealed no significant interaction of Group and Type – the Initial effect held for both groups.

Table 3. Significant effects on accuracy for entire dataset, collapsed across vowels and groups.

| | B | S.E. | t | p |
|---|---|---|---|---|
| Intercept (Teachers) | 1.683 | 0.387 | 4.35 | <.001 |
| Type=Initial | −0.289 | 0.065 | −4.441 | <.001 |
| Group=Students | −0.503 | 0.114 | −4.401 | <.001 |

Before looking at the statistical analyses for each of the vowel pairs for each listener group, consider Figure 5, which provides an overview of the accura-

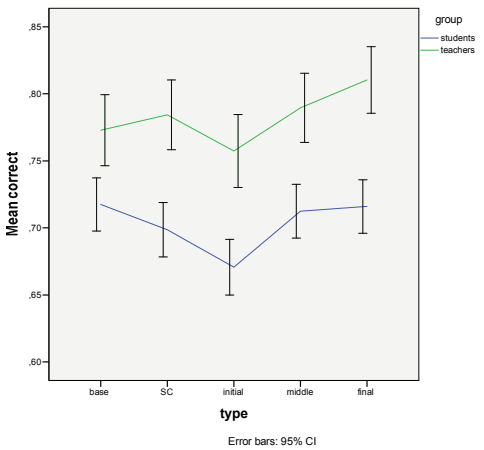


Figure 4. Mean accuracy rates for each group as a function of stimulus Type. (Colour online.)

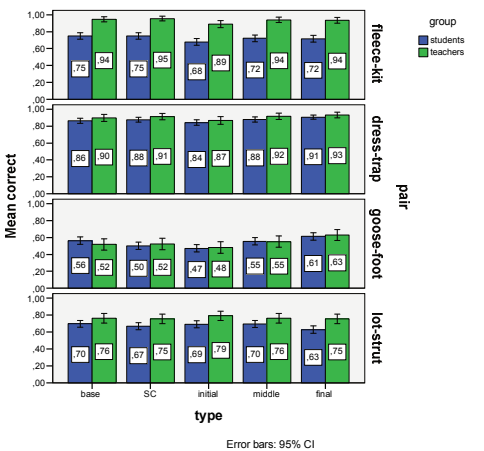


Figure 5. Overview of accuracy results for individual vowel pairs as a function of Type, sorted for Group. (Colour online.)

cy results as a function of Stimulus type for each vowel Pair. Notice that the largest group-based differences were found for the FLEECE–KIT pairs and the LOT–STRUT pairs. In the other two pairs (GOOSE–FOOT; DRESS–TRAP) accuracy rates were more or less similar between the two groups. In the figure we also

can observe that the effects of Type are largely uniform across groups. For three out of four pairs, Initial items were identified the least accurately by both groups.

### 3.5.1. FLEECE–KIT

The overall accuracy rates for the FLEECE–KIT trials were 93% for the Teachers and 72% for the students. Logistic regression analyses revealed no significant effect of Stimulus Type or Context on Teachers' accuracy. Students' responses showed no effects of Type, but a significant effect of Context by which items in the labial context were identified most accurately (Reference Level: Dorsal; B = 0.381, S.E. = .115, t = 3.309, p = .001). The effect of Context on accuracy in the Students' group is shown graphically in Figure 6.

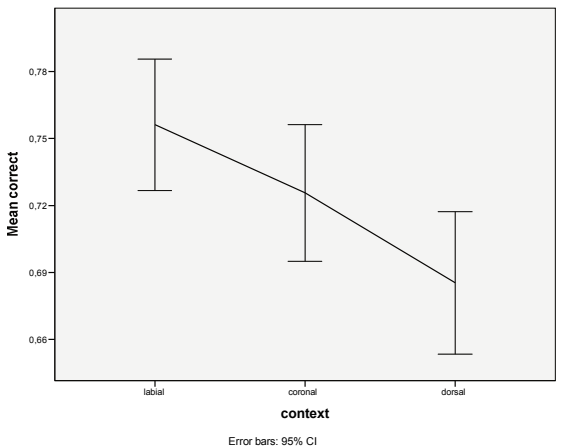


Figure 6. Mean accuracy for Students group in FLEECE–KIT trials as a function of Context. (Colour online.)

### 3.5.2. DRESS–TRAP

The overall accuracy rates for the DRESS–TRAP pairs were 91% for the Teachers and 87% for the Students. There were no significant effects of Context or Stimulus type for the Teachers. The Students showed an interaction between Type and Context by which Initial items were identified least accurately in the

Dorsal context (Reference Levels: Base-Labial; B = −0,625, S.E. = .304, t = −2.057, p = .04). This is shown graphically in Figure 7.

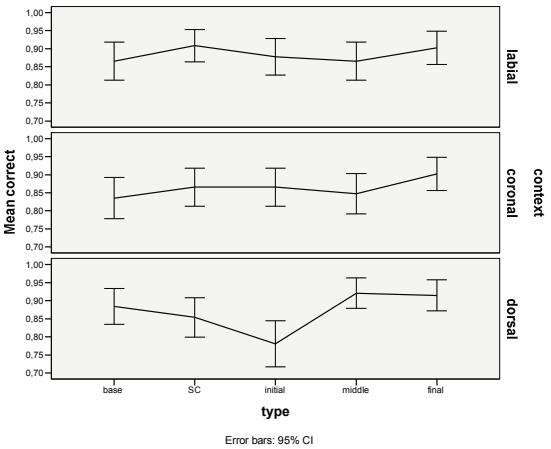


Figure 7. Mean accuracy for Students group as a function of type, sorted for Context: DRESS–TRAP trials. (Colour online.)

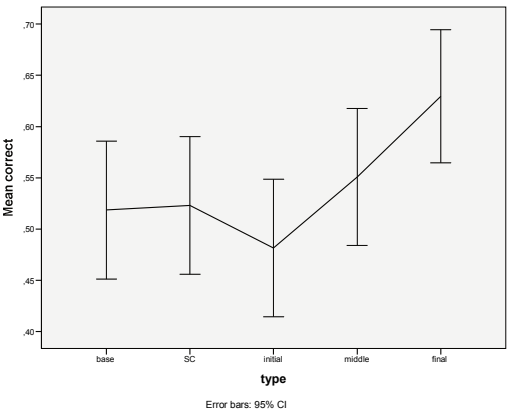


Figure 8 – Mean accuracy as a function of Type for Teachers' group: GOOSE–FOOT trials. (Colour online.)

### 3.5.3. GOOSE–FOOT

The GOOSE–FOOT pairs caused the greatest difficulty for both groups, yielding a mean accuracy rate of 54% for both groups. The Teachers showed an effect of Type by which Final tokens were identified the most accurately (Reference Level: Base; B = 0.457, S.E. = .196, t = 2.33, p = .02). This is shown in Figure 8. The slight dip in accuracy for the Initial tokens was not significant.

For the Students, Initial tokens were identified least accurately (Reference Level: Base; B = −0.368, S.E. = .128, t = −2.869, p = .004). No effects of Context were observed for either group.

### 3.5.4. LOT–STRUT

The overall accuracy rates for the LOT–STRUT pairs were 77% for the Teachers, and 68% for the Students. For the Teachers, an effect of Context was observed by which Labials were identified least accurately (Reference Level: Dorsal; B = −0.392, S.E. = .178, t = −2.202, p=.028). No effects of Type were found for the Teachers. For the Students, Final items were identified least accurately (Reference Level: Base; B = −0.320, S.E. .138, t = −2.326, p=.02). As with the Teachers, there was also an effect of Context by which labial items were least accurately identified (Reference Level: Coronal; B= −0.323, S.E.= −.108; t = 3.000, p = .003).

## 3.6. Discussion

The results of the rhyming experiment revealed nearly uniform effects of stimulus Type on identification accuracy in the two groups, despite higher overall accuracy for the Teachers' group (Figure 4). Collapsed across groups and vowel pairs, the Initial items were identified with the lowest accuracy rate. There was also an effect for both groups of consonantal Context, by which labials were identified the most accurately (except in the STRUT–LOT pair).[4] Since Initial items were characterized by the greatest movement of all the stimulus types, and

[4] The students showed lower accuracy for labials than other contexts for this pair. This was probably due to the fact that the labial context may be expected to lower F2 on the vowel. As a consequence, the STRUT–LOT distinction, which is based partially on lip rounding in British English, may be expected to be less perceptually salient in this context.

labials were characterized by the smallest magnitude of Initial interval formant movement in the stimuli (Figures 1–3 and Table 2), it may be suggested that in this experiment, greater formant movement was associated with lower identification accuracy.

This finding appears to differ somewhat from what many authors have found in L1 studies (e.g. Strange et al. 1983; Jenkins and Strange 1999). Research employing the Silent Center paradigm to investigate vowel identification in English has found that SC tokens, which may be said to promote the perceptual weight of formant movement by forcing listeners to rely on CV and VC transitions, are typically identified most accurately. As a result, it is necessary to comment at this time about the present results in light of these earlier findings.

One possible explanation is that the present experiment used stimuli recorded by a speaker of Southern British English, while the other studies focused on North American English. Looking at the formant tracks of the stimuli provided in Figures 1–3, it is evident that there is dramatic formant movement, particularly in the coronal and dorsal contexts. The figures provided by Nearey (2013) suggest that the degree of VISC in North American varieties is of a lesser magnitude. This suggestion is compatible with our own impressions of the two varieties – i.e. that vowels in Southern British English are characterized by a greater degree of diphthongization than is found in (non-Southern) accents of North American English. Perhaps it is the case that for highly diphthongized vowels the perceptual weight of the initial portion is lessened in relation to the final portion. This would explain the effect of stimulus type from the present study, by which Initial items were identified with the lowest accuracy rate. In this connection, it may be noted that experiments with L1 listeners have compared Onset-Slope and Onset–Offset models of VISC (see Morrison 2013), and found that the Onset–Offset models are the most successful at discriminating vowel quality. These findings, along with the poor performance on the Initial items in the present experiment, suggest that both L1 and L2 listeners delay identification decisions until they hear spectral cues available in later portions of the vowel.

The relative uniformity evident in Figures 4 and 5 indicates that for the most part both groups reacted similarly to VISC in the stimuli. This finding is somewhat surprising in that we might expect the group with greater proficiency in English not to be "fooled" by the initial portion of the stimuli, which contained the most formant movement (see Table 2). It is therefore not exactly clear how the uniformity between groups with regard to the effect of stimulus type may be interpreted. One possibility is that the Students' group has acquired perceptual mechanisms for vowel identification that are similar to those of the teachers.

This interpretation highlights the desirability of running this same experiment with L1 speakers of British English. Unfortunately, a homogenous group of British listeners is difficult to find in Poznan. If the uniformity were to be found to extend to native listeners as well, then the current findings could be interpreted to mean that the fine-tuning of the perceptual system to phonetic details such as VISC is well underway in learners at the B1 level. If native listeners showed different effects of Stimulus Type, then we could make a claim that vowel acquisition is persistently difficult chore for L2 learners of English, who even at the C2 level have trouble using phonetic details such as VISC to overcome equivalence classification (Flege 1995).

The results from the individual vowel pairs, which included some cases where the inter-group uniformity did not hold, may shed some light on this question. In three out of the four vowel pairs, there is some evidence that the Teachers' group has made greater progress in using VISC for vowel identification. For the FLEECE–KIT pair, the Students, but not the Teachers showed an effect by which the labial context induced higher accuracy than other contexts. Since the labial context was characterized by a smaller degree of VISC in the stimuli, it may be said that the Students group had greater difficulty dealing with the movement in the other contexts, while the Teachers had overcome this difficulty. A similar claim could be made about the Context–Type interaction for the Students (but not the Teachers) for the DRESS–TRAP pair, in which dorsal items with the greatest degree of VISC led to more identification errors. In the GOOSE–FOOT pair we found that the effect of the Initial tokens was significant for the Students but not the Teachers. In other words, hearing only the initial portion of these vowels caused greater problems for the Students than the Teachers, who may have been able to extrapolate the formant frequencies missing from the later part of the vowel.[5] The STRUT–LOT pair also showed one non-uniform effect of stimulus type by which the Students but not the Teachers were least accurate with the Final tokens. In sum, the overall picture is one of uniformity, but for some vowel pairs we may indeed witness effects of proficiency on the use of dynamic information for vowel perception.

Another striking aspect of our results is that the different pairs induced differences in overall accuracy both within groups and across groups. The largest differences in accuracy between the groups were found for the FLEECE–KIT pair followed by the LOT–STRUT pair, while the other two pairs yielded equally high

[5] It should also be mentioned that the spellings for this pair are ambiguous, which may have contributed to the overall low accuracy rates for both groups. We are grateful to a reviewer for pointing this out.

(DRESS–TRAP) and low (GOOSE–FOOT) accuracy rates in the two groups. Clearly, more L1 interference persisted in the GOOSE–FOOT pair, since even the Teachers performed with just over chance-level accuracy. Since the functional load of this pair is significantly smaller in English than for the other pairs, it may be suggested that the ability of use phonetic details such as VISC for L2 perception may be a function of input frequency. This would be in line with usage-based or exemplar approaches to phonology (e.g. Johnson 1997; Bybee 2001), for which frequency effects are crucial for the formation of phonological categories. At the same time, however, the hypothesis underlying this research is that the degree of VISC in a given language is a systemic feature, suggesting that it may be encoded by means of an abstract phonological category. In what follows, we will consider the phonological origins of VISC and possible interactions with the type of frequency effect that may underlie the findings of the current experiment.

## 4. General discussion

The phonologization of differing degrees of VISC across languages is a product of an ambiguity built into the Onset Prominence representational framework (OP; Schwartz 2013, 2016). In the OP environment, there are multiple ways of encoding the relationship between phonological structure and the acoustic signal. One such ambiguity involves the initial portion of vowels, which are acoustically vocalic but typically contain cues to the identity of the preceding consonant (e.g. Wright 2004). In the OP environment, this is represented by the Vocalic Onset (VO) node of structure. Relevant representations are shown in Figure 9. Trees (a) and (b) on the left illustrate the consonantal parse of VO, which is joined into the same hierarchical structure as Closure and Noise nodes associated with obstruents. On the right in trees (c) and (d), we see a vocalic parse of VO. Systems such as the one on the left are normally associated with a greater degree of VISC, since the consonantal representation is allowed to occupy a greater proportion of vowel duration. In other words, in such systems it may be said that consonants "intrude" into the structural space of vowels, resulting in greater consonant-vowel interactions and more formant movement. For Polish learners of English, we assume that the acquisition process involves mastering a consonantal VO system (trees a and b) after starting with Vocalic VO system (trees c and d) in L1. Thus, proficient L2 users must become attuned to the greater degree of consonant-induced formant movement found in English.

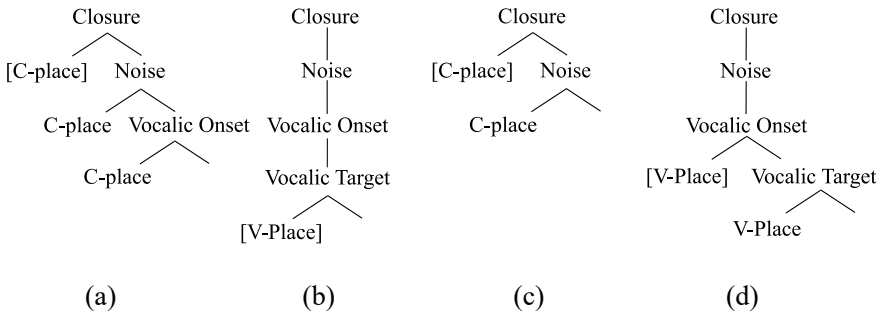


Figure 9. Consonantal (trees a and b) vs. vocalic (trees c and d) parses of VO in the OP framework. A greater degree of VISC is expected for the consonantal parse posited for English.

The VO parses shown in Figure 9 are essentially a function of prosodic organization, which governs a range of phonological phenomena that often differ systematically across languages (for details, see Schwartz 2016). At the same time, differences in vowel quality are related to the realization of melodic features that attach to OP trees. From this perspective, acquisition of L2 vowels may involve the reconfiguration of melodic specifications within the structures of the given "segments". For example, accurate production and identification of the GOOSE vowel is possible only after learners become aware that the English vowel "target" is located in the later portion of the vowel, i.e. under the Vocalic Target node, rather than under VO as in Polish. The apparent effects of frequency on identification accuracy observed in the present study, by which GOOSE–FOOT pairs were identified much less robustly than the other vocalic pairs, are perfectly compatible with this phonological perspective. Briefly stated, it may be argued that more common vowels are more conducive to the prosodic reorganization that is essential for successful acquisition. GOOSE–FOOT pairs lag behind others in acquisition. Since these pairs are heard less frequently, Polish learners are slower to associate the vowel's identity with a later portion of its duration, as suggested by systems with consonantal VO affiliation shown in Figure 9.

By offering phonetically refined representations that split vowel "segments" into multiple prosodic positions, the OP framework offers an area of compatibility between exemplar models incorporating frequency effects, and more abstract aspects of phonological representation. The key assumption of exemplar models is that phonological categories emerge on the basis of language input, which is of course rich in phonetic detail. In the OP environment, phonetic details associ-

ated with VISC are phonologized, providing insight into which emergent phonological categories may form, and how. With traditional representations, in which a vowel is a monolithic phonological unit, there are no obvious channels of communication between phonological abstraction and exemplar models.

## 5. Acknowledgement

This research is supported by a grant from the Polish National Science Centre, project nr UMO-2014/15/B/HS2/00452. We are grateful to Bartosz Brzoza and Olga Witczak for assistance with E-Prime, and to Anna Balas for providing the base recordings for our stimuli.

## REFERENCES

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

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

Blevins, J. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Boersma, P. and D. Weenink. 2017. Praat: doing phonetics by computer [Computer program]. Version 6.0.24, retrieved 6 January 2017 from http://www.praat.org/

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

Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Collins, B. and I. Mees. 2009. *Practical phonetics and phonology. A resource book for students*. London: Routledge.

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

Dukiewicz, L. and I. Sawicka. 1995. *Gramatyka współczesnego języka polskiego – Fonetyka i fonologia* [Grammar of Modern Polish – Phonetics and phonology]. Warsaw: PAN, Instytut Języka Polskiego.

Elvin, J., D. Williams and P. Escudero. 2016. "Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English". *Journal of the Acoustical Society of America* 140(1). 576–581.

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.

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: Strange, W. (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*. Timonium, MD: York Press. 233–277.

Hillenbrand, J., M. Clark and T. Nearey. 2001. "Effects of consonant environment on vowel formant patterns". *Journal of the Acoustical Society of America* 109(2). 748–763.

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

IBM Corp. 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Jacewicz, E. and R.A. Fox. 2013. "Cross-dialectal differences in dynamic formant patterns in American English vowels". In: Morrison, G. and P. Assmann (eds.), *Vowel inherent spectral change*. Berlin: Springer. 171–198.

Jekiel, M. 2010. Dynamic information for Polish and English vowels in syllable onsets and offsets. (BA 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.

Johnson, K. 1997. "Speech perception without speaker normalization: An exemplar model". In: Johnson, K. and J. Mullenix (eds.), *Talker variability in speech processing*. New York: Academic Press. 3–26.

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.

Lindblom, B. 1963. "Spectrographic study of vowel reduction". *Journal of the Acoustical Society of America* 35. 1773–1781.

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

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

Nearey, T. and P. Assmann. 1986. "Modeling the role of vowel inherent spectral change in vowel identification". *Journal of the Acoustical Society of America* 80. 1297–1308.

Ohala, J.J. 1981. "The listener as a source of sound change". In: Masek, C.S. et al. (eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society. 178–203.

Peterson, G. and I. Lehiste. 1960. "Duration of syllable nuclei in English". *Journal of the Acoustical Society of America* 32(6). 693–703.

Rogers, C.L., M. Glasbrenner, T. DeMasi and M. Bianchi. 2013. "Vowel inherent spectral change and the second language learner". In: Morrison, G and P. Assmann (eds.), *Vowel inherent spectral change*. Berlin: Springer. 231–259.
Rojczyk, A. 2011. "Overreliance on duration in nonnative vowel production and perception: The within lax vowel category contrast". In: Wrembel, M., M. Kul nad K. Dziubalska-Kołaczyk (eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010* (vol. 2). Bern: Peter Lang. 239–249.
Schwartz, G. 2013. „A representational parameter for onsetless syllables". *Journal of Linguistics* 49(3). 613–646.
Schwartz, G. 2016. „On the evolution of prosodic boundaries – paremeter settings for Polish and English". *Lingua* 171. 37–74.
Schwartz, G., G. Aperliński, J. Weckwerth and K. Kaźmierski. 2016a. "Dynamic targets in the acquisition of L2 English vowels". *Research in Language* 14(2). 181–202.
Schwartz, G., G. Aperliński, M. Jekiel and K. Malarski. 2016b. "Spectral dynamics in L1 and L2 vowel perception". *Research in Language* 14(1). 61–77.
Stevens, K.N. and A.S. House. 1963. "Perturbation of vowel articulations by consonant context". *Journal of the Acoustical Society of America* 85. 2135–2153.
Strange, W. 1989. "Evolving theories of vowel perception". *Journal of the Acoustical Society of America* 85. 2081–2087.
Strange, W., J. Jenkins and T. Johnson. 1983. "Dynamic specification of coarticulated vowels". *Journal of the Acoustical Society of America* 34. 695–705.
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.
Williams, D., J. van Leussen and P. Escudero. 2015. "Beyond North American English: Modelling vowel inherent spectral change in British English and Dutch". *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow.
Wright, R. 2004. "A review of perceptual cues and cue robustness". In: Hayes, B., R. Kirchner and D. Steriade (eds.), *Phonetically based phonology*. Cambridge: Cambridge University Press. 34–57.

**Address correspondence to:**
Geoffrey Schwartz
Faculty of English
Adam Mickiewicz University
Collegium Novum
al. Niepodległości 4
61-874 Poznań
Poland
geoff@wa.amu.edu.pl

# APPENDIX 1

Acoustic summary of base recordings for stimuli

| Vowel | Context | Mean F1(Hz) | Mean F2 (Hz) | Duration (msec) |
|---|---|---|---|---|
| DRESS | velar | 581 | 2085 | 127 |
| DRESS | labial | 632 | 1809 | 133 |
| DRESS | coronal | 593 | 1899 | 146 |
| FLEECE | velar | 308 | 2422 | 142 |
| FLEECE | labial | 302 | 2416 | 140 |
| FLEECE | coronal | 290 | 2282 | 126 |
| FOOT | velar | 400 | 1256 | 115 |
| FOOT | labial | 424 | 1058 | 117 |
| FOOT | coronal | 427 | 1397 | 120 |
| GOOSE | velar | 310 | 1456 | 116 |
| GOOSE | labial | 328 | 1064 | 113 |
| GOOSE | coronal | 338 | 1582 | 109 |
| KIT | velar | 419 | 2165 | 103 |
| KIT | labial | 460 | 2084 | 113 |
| KIT | coronal | 406 | 2156 | 133 |
| LOT | velar | 563 | 1025 | 126 |
| LOT | labial | 609 | 915 | 128 |
| LOT | coronal | 573 | 1103 | 147 |
| STRUT | velar | 713 | 1198 | 128 |
| STRUT | labial | 703 | 1098 | 149 |
| STRUT | coronal | 676 | 1313 | 140 |
| TRAP | velar | 790 | 1656 | 159 |
| TRAP | labial | 821 | 1459 | 164 |
| TRAP | coronal | 720 | 1540 | 168 |

# APPENDIX 2

## Rhyming words for perception experiment

FLEECE–KIT trials: *deep–dip*; *leak–lick*; *heat–hit*
DRESS–TRAP trials: *neck–back*; *step–tap*; *get–hat*
GOOSE–FOOT trials: *boot–put*; *Luke–look* (no pair with labial coda; extra pairs of coronal–dorsal contexts were added)
STRUT–LOT trials: *luck–rock*; *cup–top*; *nut–hot*