SPECTRAL DYNAMICS IN L1 AND L2 VOWEL PERCEPTION[*]

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

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

MATEUSZ JEKIEL
Uniwersytet Adama Mickiewicza w Poznaniu
mjekiel@wa.amu.edu.pl

KAMIL MALARSKI
Uniwersytet Adama Mickiewicza w Poznaniu
kmalarski@wa.amu.edu.pl

Abstract
This paper presents a study of L1 and L2 vowel perception by Polish learners of English. Employing the Silent Center paradigm (e.g. Strange et al. 1983), by which listeners are presented with different portions of a vowel, a force choice identification task was carried out. Due to differences in the vowel systems of the two languages, it was hypothesized that stimulus type should have minimal effects for L1 Polish vowel perception since Polish vowels are relatively stable in quality. In L2 English, depending on proficiency level, listeners were expected to adopt a more dynamic approach to vowel identification and show higher accuracy rates on the SC tokens. That is, listeners were expected to attend more to dynamic formant cues, or vowel inherent spectral change (VISC; see e.g. Morrison and Assmann 2013) in vowel perception. Results for identification accuracy for the most part were consistent with these hypotheses. Implications of VISC for the notion of cross-language phonetic similarity, crucial to models of L2 speech acquisition, are also discussed.

Keywords: Vowel perception, dynamic specification, Polish, L2 English

1. Introduction

Research into second language (L2) speech perception is characterized by a number of complexities that generally do not necessarily arise in first language (L1) perception research. These factors, of course, include the interaction between participants' L1 and L2 (and often additional foreign languages), as well as the fact that individual learners differ with regard to amount of instruction and experience, learning styles, phonetic talent and awareness, and attitude toward the target language and the learning process (for discussion, see e.g. Hansen-Edwards 2008). L1 perception research, in which the domain of study has already been 'acquired', typically does not need to address such issues and can concentrate more attention on purely acoustic and auditory considerations. As a result, there are a number of acoustic features that have been examined in L1 perception studies that have yet to be studied extensively from the perspective of cross-linguistic interaction and L2 acquisition.

Among these features we find vowel inherent spectral change (VISC; e.g. Morrison and Assmann 2013), the dynamic formant trajectories that have been observed over the course of a vowel's duration. While there is a great deal of cross-language and acquisition research examining vowel quality in terms of static positions on a two dimensional acoustic space, studies of VISC and its effects on vowel perception have been largely limited to English as an L1. It is difficult to find cross-language comparisons and acquisition research devoted to spectral dynamics (but see Jin and Liu 2013; Rogers et al. 2013).

With regard to L1 English, a growing number of studies has documented the production of VISC (e.g. Fox and Jacewicz 2009; Williams and Escudero 2014), as well as its effects on vowel perception (Strange 1989; Hillenbrand 2013). In the case of the latter, it has been established that many L1 English listeners use formant trajectories as a cue to vowel identification, and that in certain instances spectral dynamics appear to be weighted more heavily than the static position of a vowel in two-dimensional acoustic space. If we consider the fact that English has a relatively large vowel system, these findings should not be surprising. In large vowel systems, F1-F2 space is densely populated such that vowel targets show a great deal of overlap. Dynamic properties (as well as duration) allow for more robust differentiation of contrasting vowel categories. In the case of small vowel systems, however, we should expect a smaller role for spectral dynamics, since F1-F2 acoustic space more easily accommodates a smaller number of static targets. Thus, for example, a language such as Polish appears to have more stable vowel quality (see e.g. Schwartz 2015), and its listeners should have an easier time identifying static vowels in a two dimensional acoustic space.

This paper will present perceptual data from Polish listeners performing a forced-choice identification task of vowels both in L1 Polish and L2 English. In order to investigate the effects of spectral dynamics vs. static formant targets, the Silent Center (SC) paradigm was employed (Jenkins et al. 1983; Jenkins and Strange 1999). Experimental stimuli varied with respect to the portion of a given

vowel that was presented to the listener. Listeners heard tokens including only the middle portion of the vowel, tokens including only the onset of the vowel, items containing only the offset of the vowel, and a combination of the onset and offset with a silent center. The goal of the experiment is to gauge the effects of stimulus type on vowel identification. Greater effects of stimulus type, in particular increased accuracy for silent center tokens, may be attributed to a significant role of spectral dynamics in vowel perception.

The rest of this paper will proceed as follows. Section 2 will provide a brief review of vowel perception research in both L1 and L2. Section 3 will present the experiment and its results. Section 4 explores the origins of VISC and dynamic specification effects. Finally, Section 5 discusses the implications of VISC and dynamic specification for the notion of cross-language phonetic similarity that is crucial for current models of L2 speech acquisition.

2. L1 vs. L2 vowel perception

2.1. Static vs. dynamic targets in L1 vowel perception

There is a longstanding tradition in phonetics to describe vowel quality in terms of a two-dimensional chart, in which the height of a vowel is represented on the vertical axis and the front-back dimension is represented on the horizontal axis. These charts originated as impressionistic representations, which were later confirmed with the advent of acoustic analysis. Indeed, it has been shown that the acoustic dimensions of F1 and F2 encode impressionistic vowel quality more successfully than any attempted quantification of tongue position (see e.g. Ladefoged and Maddieson 1996). Early perceptual experiments established the perceptual relevance of F1 and F2 by synthesizing context-free, steady-state vowel formants, on the basis of which the perceptual identity of a vowel is described as a 'simple target' in F1-F2 space (see discussion in Strange 1989).

In further experiments, consonantal context was added to the equation, yielding some surprising results. On the production side, it was shown that these simple targets are often not reached, since co-articulation with neighboring consonants leads to target undershoot (see Strange 1989). The implications of these findings for the 'simple target' model of vowel identification are significant – how can a target that is not reached play a role in perception? Later research showed that listeners identified co-articulated vowels, in which simple targets are not reached, with surprising accuracy often exceeding that of vowels produced out of context (Strange et al. 1976). On the basis of these and other findings, Winifred Strange and colleagues formulated a theory known as the Dynamic Specification approach to vowel perception (Strange et al. 1983), in which it is hypothesized that listeners identify vowels not on the basis of static targets so much as the formant changes occurring over the course of the vowel.

Evidence for the dynamic specification approach was found using what has been referred to as the 'Silent Center' (SC) experimental paradigm (Jenkins et al. 1983; Jenkins and Strange 1999). In SC experiments, listeners are presented with the onset and offset of a vowel that have been affected by consonantal context, while the middle portion, presumably containing the 'target' vowel quality, is edited to silence. SC tokens are compared with those in which the middle portion of the vowel is preserved, as well as those preserving either the onset or the offset. Experiments found that American English listeners are more accurate in identifying SC tokens than center tokens (Jenkins et al. 1983), or those including only the onset or offset (Jenkins and Strange 1999). Thus, the most robust cues for vowel identification appear to be the formant trajectories over the course of the vowel, rather than the static formant frequencies at vowel midpoint, or any other portion of the vowel.

As mentioned earlier, however, Dynamic Specification research has been largely limited to English as an L1 – there are not many cross-language or acquisition studies testing the hypothesis. Nevertheless, considering cross-linguistic differences in vowel inventories and spectral dynamics, there is reason to believe that VISC should constitute a robust area of study for L2 research. In particular, speakers of languages with small vowel inventories should be expected to make minimal use of spectral dynamics for vowel identification. The experiments reported in this paper attempt to test this hypothesis with Polish listeners, who are also learners of English as an L2. In the meantime, however, it is necessary to provide a brief summary of vowel perception research investigating cross-language differences and L2 acquisition.

2.2. Common themes in L2 vowel perception

Much of the published research into L2 vowel perception documents the acquisition of English vowel contrasts by speakers of languages with simpler vowel systems. In particular, researchers have been interested in whether learners can discriminate notorious English vowel contrasts, such as those found in pairs such as *sheep-ship*, *look-Luke*, *men-man*, and *lock-luck*. As with much of the research in L2 acquisition, of particular interest have been the effects of factors involving linguistic experience, in particular the age at which L2 learning began.

One aspect of this research that is worthy of mention is that investigators have compared the perceptual weight of different types of acoustic cues used by listeners both in L1 and L2. In practice, this has meant duration as opposed to static formant targets in F1-F2 space. An interesting finding in this regard has been that L2 learners from L1s without vowel duration contrasts make use of 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*. Likewise, Rojczyk

(2011) found that L1 Polish speakers use duration rather than spectral patterns to distinguish English /æ/ from /ɛ/, despite the fact the Polish has no duration contrasts. Bohn (1995) would attribute such findings to a type of perceptual 'desensitization'. The idea is that since the new L2 vowel sounds under study are in close spectral proximity to an L1 sound, listeners are 'desensitized' to their spectral details. In this situation, duration is seen as the only available cue to discrimination for learners, while native speakers use both spectral and durational cues. 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, along with the models it has spawned, are invaluable aspects of our understanding of L2 speech perception, they avoid a more general question. Namely, 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? While length differences can help maintain such contrasts, duration cannot be the whole story. For example, in Scottish English, the *beat-bit* contrast is based entirely on spectral properties rather than duration (McClure 1977). In other dialects, pre-fortis clipping in countless numbers of words yields 'long' vowels that are shorter than 'short' vowels - *bid* typically has a longer vowel than *beat* even though the former is phonologically short and the latter is phonologically long. Research into VISC offers perspectives to address this issue. Thus, two vowels that have similar targets in vowel space may have greatly different dynamic properties. In English, phonologically long vowels typically show movement toward the periphery (e.g. Hillenbrand 2013), while phonologically short vowels show centralization. Thus, VISC helps maintain L1 contrasts between vowels with similar 'target' positions.

For research into L2 speech acquisition, the study of VISC, which has been invoked in the study of English dialectal variation and vowel shifts, may serve as an additional parameter for defining the oft-invoked notion of cross-language phonetic 'similarity'. Vowels may or may not be similar in their static target positions, their duration, and also their formant trajectories. Since similarity is a crucial concept for current models of L2 speech acquisition (e.g. Flege 1995), it is important to document the extent to which spectral dynamics define similarity both in production and perception.

3. Experiment

In this section we describe a pilot perception experiment run with Polish learners of English identifying vowels both in their L1 and their L2. The purpose of the experiment is to investigate the degree to which dynamic specification plays a

role in Polish vowel perception, and the extent to which exposure to English affects listeners' reliance on dynamic spectral cues in vowel identification. To our knowledge, with the exception of Jekiel (2010), this question has not been the subject of systematic experimental study.

3.1. Participants

Forty native speakers of Polish took part in the experiment, divided into two groups of twenty. One group was comprised of first year students in the Russian and English program at Adam Mickiewicz University in Poznan (UAM). That is, they majored in Russian, with English as a minor specialization. The English proficiency level of these students was estimated at B1 according to the Common European Framework for Languages. The other group was made up of advanced students in the Faculty of English at UAM, whose proficiency level in English was C1 or C2. Proficiency level (First Year – Advanced) thus comprised an independent variable, with the Advanced group having achieved higher proficiency in English, and most importantly, having completed intensive training in English phonetics.

3.2. Materials

Stimuli were recorded in an anechoic chamber at the Faculty of English at Adam Mickiewicz University in Poznań. For the L1 Polish part of the experiment, stimuli were taken from two native speakers' recordings of /bVt/ sequences containing each of the six Polish oral vowels /i ɨ ɛ a o u/. The recordings were then edited to establish four stimulus conditions of interest. In each stimulus type, different portions of the vowel were either included or left silent. Parts of the silent portions were shortened slightly to ensure more natural sounding stimuli. The stimulus types, including the portions presented and silenced, are summarized in Table 1.

Table 1. Stimulus types used in perception experiment

Stimulus Type	Description	Notes
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 of 50% of vowel duration

In the experiment, each vowel was paired with two 'incorrect' choices, one as the left option on the slide, one on the right, to counterbalance for participants' handedness. In the Polish part of the experiment there was a total of 96 trials (6 vowels*4 conditions*2 speakers*2 pairs). With 95 participants, this produced a total of 4320 Polish responses.

The English part of the experiment was concentrated on two contrasts that have been observed to be difficult for Polish learners, /i/ vs. /ɪ/ and /e/ vs. /æ/. Stimuli were taken from recordings of two native speakers of British English producing the pairs *sat-set*, *bat-bet*, *feet-fit*, and *sheep-ship*. The stimulus conditions were the same as in the Polish portion of the experiment.

The four stimulus conditions considered in this paper have implications for vowel perception in terms of the following question. Which portion or portions of a vowel are most important for listeners in vowel identification? Under the assumptions of the traditional 'simple target' model, we would expect the middle portion to be the most important, since presumably it is at vowel midpoint where the F1 and F2 values most closely resemble canonical targets. Under the 'dynamic specification' approach, formant trajectories, which should be most reliably recoverable in SC tokens, should play a dominant role in perception.

3.3. Procedure and analysis

The experiment was comprised of a two-alternative forced-choice identification task implemented in E-Prime at the Language and Communication Laboratory at the Faculty of English at UAM. In each trial, two choices were presented on a slide accompanied by an audio file. The participants used the keyboard to enter their response. They were instructed to do so as quickly as possible. E-Prime recorded accuracy and response time.[1] The experiment started with the L1 Polish trials, after which the same procedure was carried out for L2 English. Before each block, participants received instructions and 5 practice trials in the language corresponding to the block. The order of presentation of the trials in each block was randomized.

The results of the experiment were analyzed using the SPSS statistical package. For accuracy, Generalized Linear Mixed Models, with a logit transform to the binary target variable of Correct are reported. For response time (RT), Linear Mixed Models are reported. Fixed factors included Stimulus Type and Learner Group, while Participants were included as a random factor.

---

[1]   As in other studies (Volín et al. 2012), responses classified as false alarms and hesitations were excluded. The thresholds for these categories for this study were set at 150ms and 1500ms. A total of 8.7% of the responses were excluded, leaving 5695 responses included in the analizie.

3.4. Hypotheses

On the basis of cross linguistic differences in spectral dynamics, we may formulate two basic research hypotheses.

H1. Since Polish vowels are relatively pure in quality, it should make little difference for perception which portion of the vowel the listeners are presented with. This general claim leads to two sub-hypotheses for L1 Polish.

- H1a. There should be minimal effects of stimulus type on identification and response time.
- H1b. On the basis of exposure to English spectral dynamics, L2>L1 influence in the Advanced group should lead to increased accuracy on SC tokens

H2. As a result of increased exposure to spectral dynamics, the Advanced group should show effects of stimulus type on vowel identification in L2 English. These effects may be expected to be manifest in two subhypotheses.

- H2a. The Advanced group should be more accurate (and faster) on initial tokens, since they are less likely to be 'fooled' that the initial portion of the vowel contains the target
- H2b. The Advanced group should be more accurate (and faster) on Silent Center tokens since they have more experience with spectral dynamics
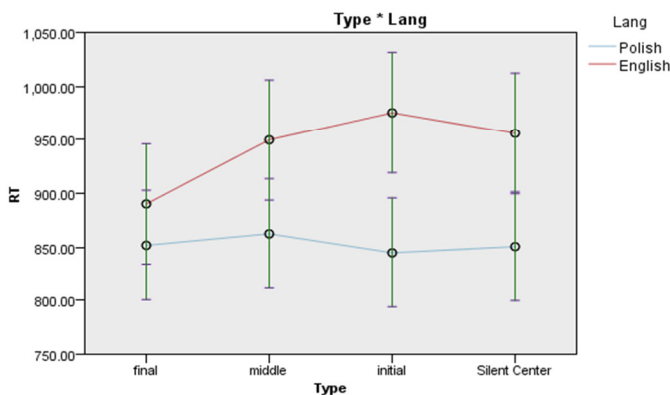
3.5. Results

The first set of results we present covers both groups of participants to look at the effect of stimulus type in L1 Polish as opposed to L2 English. In Figure 1, in which we see the overall accuracy rate was higher in L1 Polish, as might be expected, regardless of stimulus type. In L1 Polish, there was no effect of stimulus type on accuracy (p=.453). In L2 English, however, stimulus type did have a significant effect (p=.002).
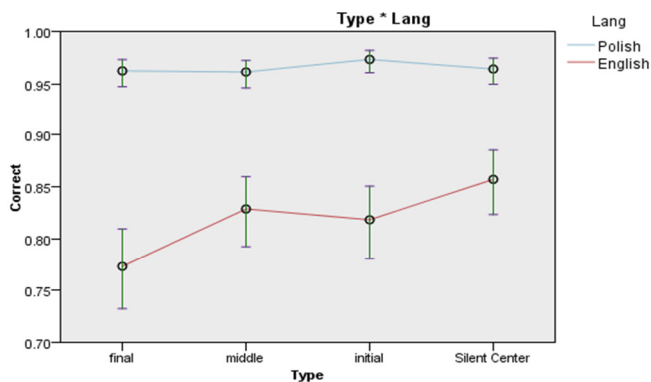
Figure 1. Identification accuracy for both groups combined as a function of stimulus type. Error bars show 95% confidence intervals.

Reaction time results for the two groups combined are given in Figure 2. As expected, responses were quicker in L1 Polish (p<.001), but there was no effect of Stimulus Type on RT (p=.826). In L2 English, there was aneffect of Stimulus Type on RT approached significance (p=.024): the Final items were identified more quickly than the other types.


Figure 2. RT for both combined groups as a function of stimulus type

The next set of results to be presented breaks down the participants into the First Year and Advanced groups. Figure 3 shows a comparison of the two groups' accuracy as a function of stimulus type overall, combining both L1 and L2. The Advanced group was more accurate in the initial (p=.024) and SC (p=.001) conditions, but not in the final and middle conditions (p>.05).
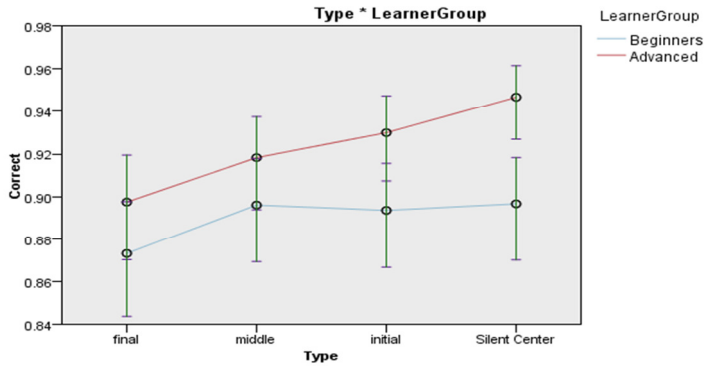
Figure 3. Accuracy for L1 and L2 combined as a function of stimulus type

Figure 4 shows accuracy in L1 Polish as a function of stimulus type. The only significant effect was in the case of the Silent Center tokens, in which the Advanced group was more accurate (p=.033)
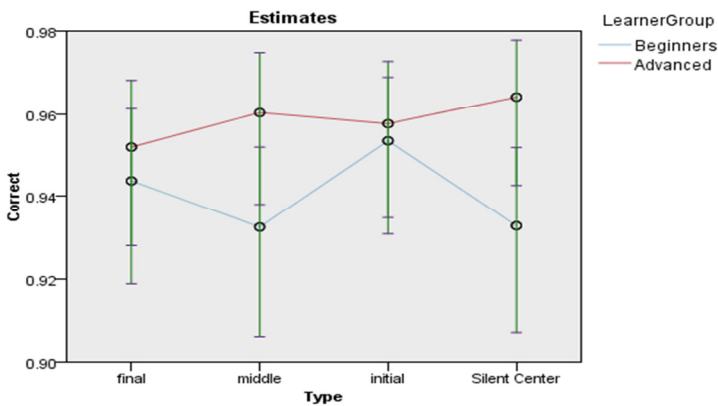
Figure 4. L1 Polish accuracy

Figure 5 shows accuracy in L2 English. The Advanced group was more accurate for the Initial (p=.023) and SC tokens (p=.043), but not for the Middle and Final items (p>.05).
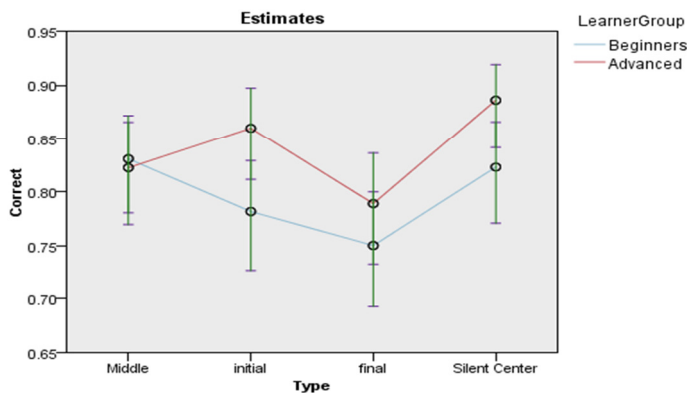
Figure 5. L2 English accuracy

Finally, Figure 6 shows RTs in L2 English.[2] For all stimulus types the Advanced group was faster (p<.05). Interestingly, across both groups, the Final tokens induced the fastest responses (p=.037), even though they also induced the highest error rates in identification.
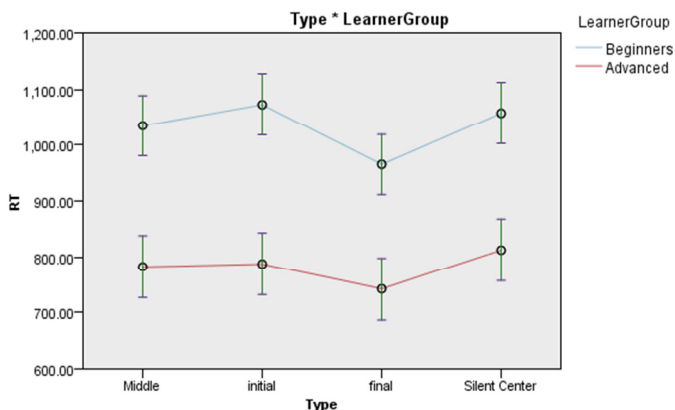

Figure 6. L2 English response times

3.6. Discussion

Our experiment examined both identification accuracy and response time (RT) in L1 Polish and L2 English by Polish learners of English. In L1 Polish, no effects of stimulus type on identification accuracy were observed (Figure 1), with one exception. The group of Advanced learners of English were more

---

[2]   In L1 Polish, no significant effects of type were observed, as suggested in Figure 2.

accurate on Polish SC tokens than the group of First year students (Figure 4). The overall pattern of L1 results provides support for Hypothesis 1a, i.e. that spectral dynamics play only a minimal role in Polish vowel perception. Since Polish has a relatively small vowel system, listeners have less impetus to rely on formant trajectories. Acoustic evidence suggests that formant movement in Polish is less salient than in English. Thus, Polish vowels have less robust spectral dynamics in production (Schwartz 2015), which provides a parallel to the perception results presented here. At the same time, the finding that SC tokens contributed to greater L1 accuracy in the Advanced group (Figure 4) suggests that C-level students may exhibit L2 influence on their L1 perception, providing support for Hypothesis 1b. Since higher accuracy on SC tokens suggests a role for dynamic specification in English vowel perception, the result indicates that greater exposure to L2 may have impacted how these listeners hear L1 Polish. L1 Polish RT results (Figure 2) are also compatible with Hypothesis 1a, since stimulus type did not have an effect.

Hypothesis 2 concerned the role of spectral dynamics in participants' identification of L2 English vowels. Since the listeners in our experiment had already achieved a certain level of competence in L2 English (B1 for the First Year group, C1-C2 for the Advanced group), it was hypothesized that vowel identification, especially in the case of the Advanced group, would resemble that of native listeners. Thus, it was expected that stimulus type would play a role in vowel perception. In particular, Hypothesis 2a predicted that the Advanced group should show greater accuracy on SC tokens having acquired the dynamic cues used by native listeners. Hypothesis 2b, similarly, predicted higher accuracy for the Advanced group on Initial tokens, since the more proficient learners may be less likely to be 'fooled' by the formant frequencies in the early part of the vowel. Both of these hypotheses were supported by the data. The fact that no between-group accuracy differences were found on the Middle and Final tokens is also compatible with Hypothesis 2. Dynamic specification effects are dependent on hearing formant trajectories, which are most easily identifiable as changes from the initial portion of the vowel onwards. The initial portion of the vowel was absent from the Middle and Final tokens, so the more proficient Learners did not receive the proper starting points from which to calculate formant trajectories. With regard to L2 response time results (Figure 6), as expected the Advanced group was faster. One unexpected finding was that the Final tokens, which were identified least accurately by both groups (Figure 5), induced the quickest responses.

Beyond Jekiel (2010), the present experiment represents one of the first attempts at an experimental study on the role of spectral dynamics in Polish vowel perception, as well as in Polish listeners' identification of L2 English vowels. In the context of the current state of research into Vowel Inherent Spectral Change and its perceptual consequences, which is for the most part limited to English as an L1, the experiment described in this paper may therefore be seen as a worthwhile contribution to VISC research. More generally, the

study of VISC may have important implications for the study of L2 speech acquisition that have yet to be thoroughly explored. These implications stem from the potential role of spectral dynamics in defining cross-language phonetic similarity, a concept that this crucial for current models of L2 speech. In what follows, we will take up theoretical issues underlying these considerations, starting with the origins of VISC in the phonetic interaction between vowels and neighbouring consonants.

4. The origins of VISC and its perceptual implications

A primary role in the emergence of formant dynamics over the course of vowel duration may be attributed to the co-articulatory effects of neighbouring consonants. In the case of pre-vocalic consonants, CV co-articulation creates transitional formant movement that typically lasts about 20% of the duration of the vowel.[3] Perceptual experiments with L1 English speakers have established that listeners attend to these transitional cues in their identification of the place of articulation of the consonant (see e.g. Wright 2004). In other words, the consonant is identified in large measure on the basis of acoustic information located on the vowel.

   If 20% is set as a reference point for the duration of CV transitions, we may consider consequences for vowel quality when that transition is pronounced more quickly, or more slowly. When the transition is produced more slowly, the co-articulatory effects of the consonant produce formant movement that occupies a greater percentage of the vowel, extending to vowel midpoint or even further. When dynamic formant patterns occupy a greater portion of the vowel, we should expect that listeners should be more likely to attribute the movement to an inherent part of the vowel representation, rather the contextual effects of the neighbouring consonant. That is, rather than using the formant movement to perceptually 'reconstruct' the consonant (cf. Ohala 1981), listeners reinterpret it as VISC. In this way, the role of VISC is phonologized, so what starts out as a co-articulatory effect of consonants shifts its affiliation to become a feature inherent to vowels.

   Since different consonant places of articulation produce different formant patterns, it may be hypothesized that dynamic specification originates with the most common consonantal context, typically coronal, and then extends analogically to all instances of a given vowel. That is, formant trajectories in dynamic specification may be regularized to the effects of a single consonantal context.

   When CV (and VC) transitions are produced more quickly, the effects on the vowel system are quite different. More rapid transitions mean that formant

---

[3]   Many studies of VISC production ignore the first and last 20% of vowel duration, attributing it to the effects of consonantal context (e.g. Fox and Jacewicz 2009).

targets are reached earlier in the vowel, and may also be expected to occupy a larger percentage of vowel duration. This results in vowels that are more pure (less diphthongal) in quality. Looking at this question from a different perspective, it may be said that in languages with 'pure' vowels, CV transitions must be rapid, since slower CV transitions are associated with diphthongization. Thus, more rapid CV transitions entail a lesser degree of VISC in production, which would have us expect a smaller role for dynamic specification in perception.[4]

A comparison of English and Polish reveals that English is a language with slower transitions between consonant and vowel (and vowel and consonant), while Polish is the opposite, with more rapid transitions. Evidence for this generalization may be found in cross-language studies of consonant place perception. More rapid transitions should be associated with a lesser reliance on contextual effects for consonant identification and greater relative weight of consonant noise bursts for perception. This prediction was confirmed for CV contexts in studies by Schwartz and Aperliński (2014) and Aperliński and Schwartz (2015), which revealed that Polish listeners place greater perceptual weight on noise cues than L1 English listeners. With regard to VC contexts, this notion is compatible with the fact that stops in coda position in Polish must always be produced with audible release (except in homorganic clusters), while English coda stops are often left unreleased. In this connection, Schwartz et al. (2014) found that more proficient Polish users of English produce longer VC formant transitions than beginners, and they are more likely to produce unreleased stops.

The fundamental difference between Polish and English with regard to CV and VC transitions may be attributable to differences in articulatory setting. Święciński (2004) provides a comparison of articulatory settings in Polish and English for the purposes of L2 pronunciation teaching. He notes that Polish is characterized by a greater degree of muscular tension in the supra-laryngeal areas of the vocal tract. Greater muscular tension is associated with more rapid articulator movements, accounting for the acoustic and perceptual patterns discussed above.

## 5. Final Remarks - VISC and cross-linguistic similarity

A question that must be addressed at this point is how the consonant-vowel interactions that result in cross-language differences in VISC and dynamic specification may be expressed in terms of the notion of 'similarity', which is

---

[4]   Note that rapid CV transitions do not exclude the possibility of more stable coarticulatory effects on static vowel targets. Thus, for example, the Polish word *ciocia* aunt, has an /o/ surrounded by pre-palatal consonants that is typically fronted. However, a fronted target does not necessarily entail longer transitions.

crucial for current models of L2 speech acquisition. In this connection, consider Figure 7, which shows spectrogram displays of the English word *deed*. The token on the left is a native-like realization produced by an L1 Polish professor of English. The token on the right is Polish-accented production produced by a first-year student. Notice that there is a very large difference between the two tokens in the F2 trajectory, with longer F2 transitions off of and onto the neighbouring consonants in the professor's pronunciation. This of course, is indicative of a diphthongal realization of this vowel that is common in many native varieties.

A heavily accented Polish English production of *deed* sounds a lot like *did*. The average F1-F2 over the first 25% of the native-like vowel in Figure 7 also resembles the F2 of *did*. Thus, it may be suggested that Polish listeners, in forming new categories for the L2 vowel, may do so on the basis of what they hear in the initial portion of the vowel. That is, since there are minimal dynamic specification effects in Polish, learners may simply base their percept on the average formant values from the beginning of the vowel, without attending to the formant trajectory. These considerations, along with the data presented in this paper, suggest that a fuller understanding of 'similarity' in cross-language vowel perception must include a determination of which portion of an L2 vowel is similar to which portion of an L1 vowel. That is, in a language pair like Polish and English, VISC and dynamic specification should play a role in establishing what is similar to what.
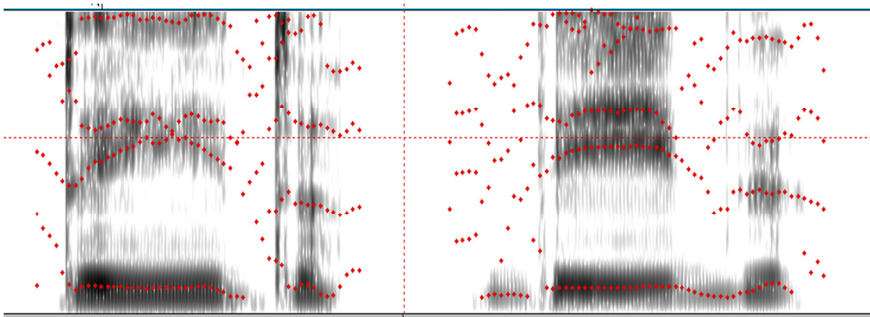
Figure 7. Spectrograms of *deed* produced by a L1 Polish professor with native-like pronunciation (left) and a first year student with a heavy Polish accent (right).

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*.

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.

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.

Jekiel, M. 2001. Dynamic information for Polish and English vowels in syllable onsets and offsets. Unpublished B.A. thesis, Adam Mickiewicz University in Poznań.

Hansen-Edwards, J.G. 2008. Social factors and variation in production in L2 phonology. In J. G. Hansen Edwards and M. Zampini (eds.), *Phonology and second language acquisition*, 251-279. Amsterdam: John Benjamins.

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.

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

Jenkins, J.J., Strange, W. and T. R. Edman. 1983. Identification of vowels in 'vowelless' syllables. Perception and Psychophysics 34, 441-450.

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

Ladefoged, P. and I. Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

McClure, J. D. 1977. Vowel duration in a Scottish accent. Journal of the International Phonetic Association 7, 10-16.

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

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.

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.

Strange W., R. Verbrugge, D. P. Shankweiler and T. R. Edman. 1976. Consonantal environment specifies vowel identity. Journal of the Acoustical Society of America 60, 213-224.

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.

Volín, J., M. Uhrinova and R. Skarnitzl. 2012. The effect of word-initial glottalization on word monitoring in Slovak speakers of English. Research in Language 10, 173-181.

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.

Wright, R. 2004. Perceptual cue robustness and phonotactic constraints. In B. Hayes, R. Kirchner and D. Steriade (eds.), *Phonetically Based Phonology*, 34-57. Cambridge: Cambridge University Press.